Welfare Economics at the Extensive Margin
Giving Gorman Polar Consumers Some Latitude

Daniel McFadden[1]

W. M. Gorman Commemorative Conference, London, June 11, 2004
Public Economic Theory Conference, Beijing, August 27, 2004
Nobel Conference, Lindau, September 3, 2004
Entretiens Jacques-Cartier: Transports, Montreal, October 7, 2004

ABSTRACT: Applied welfare economics is greatly facilitated when the preference field of consumers can be restricted so that compensating and equivalent variations have computationally tractable forms that can be estimated from market data. The leading class of preferences of this type is the Gorman polar form, particularly with "parallel Engle curve" restrictions that permit aggregation of preferences into a community preference field. This paper reviews the origins and properties of the Gorman polar form, and its use in welfare economics, and explores the extension of this form to problems of consumer choice in hedonic or physical space. An assumption that Engle curves are parallel across locations leads to simplifications in the description of demand and the welfare calculus that use "locationally representative" Gorman consumers. The paper considers two applications. The first compares the deadweight losses from Ramsey regulation and self-regulation of an industry with Cournot retailers that utilize a common network resource such as a transportation or communication network. The second analyzes consumers' willingness-to-pay to remediate an environmental hazard that has a spatial distribution around a point source, or to evaluate the spatial welfare effects of a transportation network improvement. The applications illustrate the usefulness of welfare analysis employing Gorman preference fields that are given sufficient latitude so that they approximate the true consumer preference field.

KEYWORDS: Gorman_polar_form, compensating_variation, equivalent_variation, representative_consumer, Ramsey_pricing, hedonic_regression

JEL CLASSIFICATION: D4, D60, L13, L50, R10, R31, Q25

---

[1]Department of Economics, University of California, Berkeley, mcfadden@econ.berkeley.edu, or http://emlab.berkeley.edu/~mcfadden. This research was supported by the E, Morris Cox Endowment. I am deeply indebted to Marius Schwartz for comments on Section 4 of this paper, and to James Heckman and Rosa Matzkin for comments on Sections 6 and 7. I retain sole responsibility for errors.

Welfare Economics at the Extensive Margin
Giving Gorman Polar Consumers Some Latitude

Daniel McFadden

1. INTRODUCTION

One of the ironies of science is that the most enduring legacy of a lifetime of research is often encapsulated in some simple formula, too useful to be called a cartoon, but nevertheless masking much of the subtlety of the thought behind it. This is the case for Einstein's sublime $E = mc^2$, and it is also the case for Terence Gorman's useful and revealing polar form for indirect utility, $u = (y - b(p))/a(p)$. This paper is an abbreviated survey of the origins of this form, its current role in welfare economics, and its extension to the analysis of distributed environmental or network effects on consumer welfare in hedonic or physical space, where preference heterogeneity and behavior at extensive margins determine consumer outcomes.

When Terence Gorman began his research career in the 1950's, consumer theory and welfare economics were already mature subjects. Hicks (1939) and Samuelson (1947) had worked out the properties of individual demand functions using the methods of constrained optimization. Samuelson's characterization of income and substitution effects, using the envelope theorem and bordered hessians, stands as a high point in the use of calculus in economics. The Marshallian formulation of consumer surplus had been sharpened into modern form by Hicks. However, empirical demand analysis was in its infancy, primarily because in those days of the handwritten ledger and mechanical calculator, data was limited and analysis intractable beyond linear regression in a few variables. This changed rapidly with the dawning of the computer age at the end of the 1950's, and the arrival of the first extensive data on individual demand behavior. Terence was caught up in the search for tractable, theoretically-consistent analytic methods for handling individual consumer data, as were many of us who came of age in that era. Terence was not a "numbers guy" like Zvi Griliches or Richard Stone, but he had a great eye for the theoretical simplifications that translate into useful empirical forms. With today's technology, computational feasibility is well down the list of considerations in empirical consumer theory, and attention to parsimonious parametric functional forms seems quaint, but even so, Terence's methods live on and form the core of the theory underlying empirical demand studies.

A great theoretical innovation in microeconomic analysis in the 1950's, duality theory, provided the ingredients for much of Terence's work, and others of us as well. Some elements of this theory, notably the indirect utility function and its relation to market demand functions, and the unit cost function and its relation to unit factor demands, were known earlier (Roy, 1943; Hotelling, 1935,1938; Samuelson, 1947). However, the full power of duality theory came to economics from Fenchel's unpublished Princeton University lecture notes on convexity, through Ron Shephard's 1953 book on cost functions and Lionel McKenzie's 1957 paper on demand analysis. I learned about these methods in 1961 from Mark Nerlove and Hirofumi Uzawa, who saw their potential for both theoretical and empirical analysis. Convexity, duality, and the recovery of cost functions from differential characterizations of demand became the subject of my thesis, and led to some interaction between Terence and me.[2] Terence had a deep understanding of duality, but rarely used dual arguments explicitly in his papers. Looking back, this made them less transparent than they might have been. A number of his most significant contributions, on separability, aggregation, and the polar form itself, had their broadest impact after translations by Deaton and Muellbauer (1980) and Blackorby, Boyce, and Russell (1978) that exploited the dual simplifications.

2. THE WELFARE CALCULUS

The dual formulation of consumer demand analysis and the welfare calculus is a finished subject well covered in textbooks; see Varian (1992, Chap. 7, 10) and Mas-Colell, Whinston, and Green (1995, Chap. 3E,F,G,I). For notation, I give a capsule summary. Suppose a consumer has a continuous utility index U(x,z,ρ) defined on a set **X** of commodity vectors x, a set **Z** of vectors z giving the hedonic attributes of these commodities and of the environment,

[2] Terence somehow heard about my work, requested a copy of the thesis as it was being typed in 1962, and received one of the five copies ever made. That fall, while I was a post-doc at the University of Pittsburgh, Terence invited me to become a research fellow at Oxford, but my family and I were induced instead to Berkeley at a more sustainable salary of $8200 per year. This was as close as I came to becoming Terence's protege. He already had at that time the reputation of being one of the best and most inquisitive minds in economics, and I still wonder whether I would have turned out differently under his tutelage. One thing I am sure of – the interaction would have been interesting. The substantive contents of my thesis were subsequently published in McFadden (1963, 1966, 1978abc). Many of its mathematical results are obtained and stated more generally in Rockafellar (1970).

and a set **R** of vectors ρ characterizing tastes.[3] Suppose the consumer seeks to maximize utility subject to a budget constraint $y \geq p \cdot x$, where p is an n-dimensional vector of non-negative prices in a cone **P**, and y is an income level higher than the minimum necessary to make at least one vector in **X** affordable.[4] Suppose the configuration of utility and the consumption set **X** are such that in the relevant range, local non-satiation holds; e.g., at least one commodity is available in continuous amounts and always desired. The reason for including z as an argument in U is to permit welfare analysis of non-market changes in product attributes or changes in levels of non-market goods, while ρ is included for later analysis of the effects of taste heterogeneity. Assume that **Z** and **R** are compact topological spaces, and that **X** is a closed subset, bounded below, of a finite-dimensional Euclidean space.[5] Let **U** denote the range of U.[6] In general, we do not require that **X** be a convex set, or that preferences be convex (i.e., we do not require that U be a quasi-concave function). Define the *expenditure function*

(1) $y = M(p,u,z,\rho) = \min_{x \in \mathbf{X}}\{p \cdot x \mid U(x,z,\rho) \geq u\}$,

and the *Hicksian (compensated) demand function*

---

[3]The existence of a continuous utility index for preferences is useful for welfare analysis, but more than is needed for most duality and demand analysis purposes. Note that we assume the utility index is jointly continuous in (x,z) and tastes (preferences) ρ. The following notational translation of McFadden and Train (2000), Lemma 1, gives a preference continuity property that is sufficient for the existence of such a continuous utility index: Suppose consumers with tastes defined by points ρ in a compact topological space **R** have preferences over objects (x,z) in a compact topological space **X**×**Z**, with $(x',z') \succeq_\rho (x'' z'')$ meaning (x',z') is at least as good as $(x'' z'')$ for a consumer with tastes ρ. Suppose $\succeq_\rho$ is complete and transitive, and has the continuity property that if a sequence $(x'^k, z'^k, x''^k, z''^k, \rho^k)$ converges to a limit $(x'^0, z'^0, x''^0, z''^0, \rho^0)$ and satisfies $(x'^k,z'^k) \succeq_{\rho^k} (x''^k,z''^k)$, then $(x'^0,z'^0) \succeq_{\rho^0} (x''^0,z''^0)$. Then there exists a utility function U(x,z,ρ), continuous in its arguments, that represents $\succeq_\rho$ for $\rho \in \mathbf{R}$.

[4]If **X** has a bliss point, then y is restricted to be less than the cost of the bliss point. It is possible to assume that **X** is a (continuous) function of ρ, but we will not need this generalization.

[5]When it is useful to have **X** compact, this can be accomplished by imposing a bound that is not economically restrictive. Most of the results of duality theory continue to hold when prices p are points in a convex cone in a locally convex linear topological space **P**, and **X** is a compact subset of its conjugate space **P***. This extension is useful for applications where the consumer is making choices in physical or hedonic space.

[6]We omit the generalization that makes **U** a function of ρ.

(2) $x = H(p,u,z,\rho) = \operatorname{argmin}_{x\in\mathbf{X}}\{p\cdot x \mid U(x,z,\rho) \geq u\}$,

which is in general a upper hemicontinuous correspondence for $p \in \mathbf{P}(u,z,\rho)$ and $(u,z,\rho) \in \mathcal{U}\times\mathbf{Z}\times\mathbf{R}$, where $\mathbf{P}(u,z,\rho)$ is the cone of prices where the minimum is attained; the interior of this cone is the positive orthant, and its closure is the non-negative orthant.[7] The expenditure function is strictly increasing in u, and concave and linear homogeneous in p, and consequently almost everywhere twice continuously differentiable in p with symmetric second derivatives.[8] Its epigraph

(3) $\mathbf{A}(u,z,\rho) = \{(p,y)\in\mathbb{R}^{n+1} \mid y \leq M(p,u,z,\rho)\}$

is a closed cone, and a vector x is a support of $\mathbf{A}(u,z,\rho)$ at p (i.e., $q\cdot x \geq M(q,u,z,\rho)$ for all $q \in \mathbf{P}(u,z,\rho)$), with equality for q = p, if and only if it is in the convex hull of $H(p,u,z,\rho)$.[9]

Define the *indirect utility function*

(4) $u = V(p,y,z,\rho) = \max_{x\in\mathbf{X}}\{U(x,z,\rho) \mid y \geq p\cdot x\}$

for positive p and $y > \min_{x\in\mathbf{X}} p\cdot x$, and the *market demand function*

(5) $x = D(p,y,z,\rho) = \operatorname{argmax}_{x\in\mathbf{X}}\{U(x,z,\rho) \mid y \geq p\cdot x\}$,

where in general D is a upper hemicontinuous (in p,y,z,ρ) correspondence.[10]

The expenditure function and indirect utility function satisfy the identities

---

[7]The cone $\mathbf{P}(u,z,\rho)$ is not necessarily convex.

[8]This property holds only when p is finite-dimensional.

[9]See, for example, McFadden (1966,1978abc), Diewert (1974,1982). The dual mapping from an expenditure function to a utility function is $U(x,z,\rho) = \max\{u \mid p\cdot x \geq M(p,u,z,\rho) \text{ for all } p>>0\}$. If M is the expenditure function of a quasi-concave, non-decreasing utility function U on convex $\mathbf{X}$, then the dual mapping returns U; otherwise, it returns the quasi-concave closed free disposal hull of U on the closed convex hull of $\mathbf{X}$.

[10]If V is the indirect utility function of a quasi-concave, non-decreasing utility function U on convex $\mathbf{X}$, then the dual mapping $U(x,z,\rho) = \min_p V(p,p\cdot x,z,\rho)$ returns U; otherwise, it returns the quasi-concave closed free disposal hull of U on the closed convex hull of $\mathbf{X}$.

$$(5)\qquad y \equiv M(p,V(p,y,z,\rho),z,\rho) \equiv p\cdot H(p,V(p,y,z,\rho),z,\rho)$$

$$D(p,y,z,\rho) \equiv H(p,V(p,y,z,\rho),z,\rho),$$

$$V(p,y,z,\rho) \equiv U(D(p,y,z,\rho),z,\rho).$$

Shephard's identity establishes that when M is differentiable in p,

$$(6)\qquad H(p,u,z,\rho) \equiv \nabla_p M(p,u,z,\rho),$$

while Roy's identity establishes that when V is differentiable in p and in y,

$$(7)\qquad D(p,y,z,\rho)\nabla_y V(p,y,z,\rho) \equiv -\nabla_p V(p,y,z,\rho).$$

Substituting the indirect utility function into the expenditure function gives a monotone increasing transformation that is again a utility function, now denominated in dollars and termed a *money-metric utility function*,

$$(8)\qquad u = \mu(p',z';p,y,z,\rho) \equiv M(p',V(p,y,z,\rho),z',\rho),$$

where (p',z') determine the metric and (p,y,z) determine the utility level. This function behaves like an expenditure function in p' and an indirect utility function in (p,y), and satisfies $\mu(p,z;p,y,z,\rho) \equiv y$.

Consider a change from (p',y',z') to (p'',y'',z''). The *Compensating Variation* or *Willingness-to-Pay* (WTP) for this change is the net reduction in final income that makes the consumer indifferent to the change,

$$(9)\qquad CV = \mu(p'',z'';p'',y'',z'',\rho) - \mu(p'',z'';p',y',z',\rho) \equiv y'' - \mu(p'',z'';p',y',z',\rho)$$

$$\equiv \{y'' - y'\} - \{\mu(p'',z'';p',y',z',\rho) - \mu(p'',z';p',y',z',\rho)\}$$
$$- \{\mu(p'',z';p',y',z',\rho) - \mu(p',z';p',y',z',\rho)\}.$$

The last identity decomposes the compensating variation into the net increase in money income, less the net increase in income necessary at final prices and initial utility level to offset the change in non-market attributes, less the net increase in income necessary to offset the change in prices at the initial non-market attributes and utility level. The third term can be written

$$(10)\quad \mu(p'',z';p',y',z',\rho) - \mu(p',z';p',y',z',\rho) = \int_0^1 H(p(t),u',z',\rho)\cdot\nabla_t p(t)dt,$$

the *Hicksian net consumer surplus* from the change in prices from p' to p'', where $u' = V(p',y',z',\rho)$ is the initial utility level. This integral is taken over any rectifiable path $p(t)$ from $p(0) = p'$ to $p(1) = p''$, and is independent of path.

The *Equivalent Variation* or *Willingness-to-Accept* (WTA) the change is the net addition to initial income that makes the consumer indifferent to the change,

$$(11)\quad EV = \mu(p',z';p'',y'',z'',\rho) - \mu(p',z';p',y',z',\rho) \equiv \mu(p',z';p'',y'',z'',\rho) - y'$$

$$\equiv \{y'' - y'\} - \{\mu(p'',z'';p'',y'',z'',\rho) - \mu(p'',z';p'',y'',z'',\rho)\}$$
$$- \{\mu(p'',z';p'',y'',z'',\rho) - \mu(p',z';p'',y'',z'',\rho)\}.$$

Again, the final decomposition is into the net increase in money income, less the net increase in income necessary at final prices and utility level to offset the change in non-market attributes, less the net increase in income necessary to offset the change in prices at the initial non-market attributes and final utility level, with the last term expressible as a Hicksian consumer surplus integral analogous to (10), but with Hicksian demand evaluated at the final utility level $u'' = V(p'',y'',z'',\rho)$.

Suppose a population of consumers is in general heterogeneous in tastes, income, and hedonic environment, but faces common market prices p. Let $\Psi(\rho)$ denote the distribution of tastes, which under the classical assumption of consumer sovereignty is independent of income and the hedonic environment, and invariant under changes in income, market, and non-market conditions. Let $\Gamma(y|\rho,Y)$ denote the distribution of income, conditioned on tastes and income policy parameters (e.g., per capita mean income, denoted by Y). Let $\Phi(z|\rho,y,\zeta)$ denote the distribution of hedonic environments, conditioned on tastes, income, and environmental policy parameters ζ. In applications, the z might be hedonic attributes of commodities such as durability or reliability, and the ζ might be regulations on product quality. Alternately, the z might be environmental attributes such as air pollution level or proximity to a hazardous waste site, and the ζ might be environmental regulations. In some cases, the z are exogenous to the consumer, and thus independent of income and tastes. For example, a product attribute such as durability may be uniform for all consumers. In other cases, the z are an *endogenous* consequence of consumer choice, such as residential location in response to air pollution levels, and thus have a distribution that depends on income and tastes. A satisfactory model for WTP in the presence of endogenously determined environmental attributes requires specification of the structure of supply as well as demand, and determination of an equilibrium allocation in both market goods and the non-market environments. WTP is then defined on an equilibrium trajectory from old to new environmental, income, and price management policies.[11]

Mean WTP in the population is obtained by integrating (9) with respect to the three distributions Φ, Γ, and Ψ. Even without the weighting for income and environmental circumstance that might be dictated by a social welfare function, this mean WTP is potentially difficult to obtain, requiring recovery of individual money metric utility from individual demands, including tastes for non-market goods, and identification of the conditional distributions taking into account the possible correlation of income and tastes and the possible endogeneity of z. One approach to managing this problem is to restrict preference fields in ways that allow aggregation over consumers, concentration on subsets of commodities or commodity aggregates, and simplification or sparse parameterization of money-metric utility. Gorman's polar form, and the families of empirical demand systems built upon it, have played a central role in this approach to identifying consumer welfare effects. A second, overlapping, approach has been to start from empirical descriptions of market demands or market equilibrium, and solve the integrability problem to reconstruct indirect utility functions and calculate or bound Hicksian consumer surplus measures. This approach has been used by Willig (1976), Dubin and McFadden (1984), and Hausman (1970,1981) to obtain parametric estimates or bounds on WTP. The important next step of combining supply and demand and recovering structure from equilibrium observations, has been taken by Brown and Matzkin (1998) and Heckman, Matzkin,

[11] In the terminology of the statistical and econometric literature on treatment effects, the final state in the welfare comparison is a fully consistent equilibrium counterfactual.

and Neshelm (2003ab). Their nonparametric indirect utility estimates permit estimation of WTP. In this paper, I concentrate on the first approach, but conclude with comments on the use of the two approaches in tandem.

3. THE GORMAN POLAR FORM

Consumers' preferences have a *Gorman polar form* if they are characterized by indirect utility functions

(12) $u = (y - b(p,z,\rho))/a(p,z,\rho)$,

where $b(p,z,\rho)$ and $a(p,z,\rho)$ are concave, linear homogeneous, non-decreasing upper-semicontinuous functions of p.[12] This form first appears in Gorman (1953, Theorem IX), in a demonstration that parallel affine linear Engle curves are necessary and sufficient for the existence of community indifference curves; see also Chipman and Moore (1980, 1991). Gorman (1961) reformulates this preference field in terms of expenditure functions, and in the language of duality refers to (12) as the polar form of a field of direct utility functions. Blackorby, Boyce, and Russell (1978) give a detailed and useful discussion of this preference field, including a full characterization of the structure of direct utility. Summarizing, (12) inverts to the expenditure function

(13) $M(p,u,z,\rho) = b(p,z,\rho) + u\cdot a(p,z,\rho)$,

and yields the money metric utility function

[12] With these assumptions on the functions a and b, and the domain restriction $y > b(p,z,\rho)$, (12) has the properties of an indirect utility function; i.e., it is homogeneous of degree zero and quasi-convex in (p,y), increasing in y, and non-increasing in p. The quasi-convexity follows from the concavity of the expenditure function and duality, and can also be demonstrated directly: Consider (p',y') and (p'',y''), and a proper linear combination $(p^\theta,y^\theta) = \theta\,(p',y') + (1-\theta)(p'',y'')$. By homogeneity, prices and incomes can be normalized so that $a(p',z,\rho) = a(p'',z,\rho) = 1$, and the concavity of a then implies $a(p^\theta,z,\rho) \geq 1$. Then, concavity of b implies $(y^\theta - b(p^\theta,z,\rho))/a(p^\theta,z,\rho) \leq \max\{(y' - b(p',z,\rho))/a(p^\theta,z,\rho),(y'' - b(p'',z,\rho))/a(p^\theta,z,\rho)\} \leq \max\{(y' - b(p',z,\rho))/a(p',z,\rho),(y'' - b(p'',z,\rho))/a(p'',z,\rho)\}$, with the last inequality following from the positivity of $y' - b(p',z,\rho)$ and $y'' - b(p'',z,\rho)$.

(14) $\mu(p',z';p,y,z,\rho) \equiv b(p',z',\rho) + a(p',z',\rho)\cdot(y - b(p,z,\rho))/a(p,z,\rho)$,

and the Hicksian and market demand functions

(15) $H(p,u,z,\rho) = \nabla_p b(p,z,\rho) + u\cdot\nabla_p a(p,z,\rho)$,

(16) $D(p,y,z,\rho) = \nabla_p b(p,z,\rho) + [(y - b(p,z,\rho))/a(p,z,\rho)]\nabla_p \log a(p,z,\rho)$,

the latter having affine linear Engle curves.[13] Gorman polar form preferences yield simple formulas for compensating and equivalent variation, which differ only through the scale factor

(17) $CV = a(p'',z'',\rho)[(y'' - b(p'',z'',\rho))/a(p'',z'',\rho) - (y' - b(p',z',\rho))/a(p',z',\rho)]$,

$EV = a(p',z',\rho)[(y'' - b(p'',z'',\rho))/a(p'',z'',\rho) - (y' - b(p',z',\rho))/a(p',z',\rho)]$.

In a population of Gorman polar consumers, market level average demands are obtained by integrating over the joint distribution $\Phi(z|\rho,y,\zeta)\Gamma(y|\rho,Y)\Psi(\rho)$ of non-market attributes, income, and preferences. When the function $a(p,z,\rho)$ is independent of z and ρ, and the distribution $\Phi(z|\rho,y,\zeta)$ is independent of y, market level average demands have the form

(18) $\int_R D(p,y,z,\rho)\Phi(dz|\rho,y,\zeta)\Gamma(dy|\rho,Y)\Psi(d\rho) = \nabla_p B(p,\zeta) + (Y - B(p,\zeta))/a(p))\nabla_p \log a(p)$,

where Y is average income and $B(p,\zeta) = \int_R b(p,z,\rho)\Phi(z|\rho,\zeta)\Psi(\rho)$. These average demands coincide with the demands of a "representative" Gorman polar consumer with indirect utility $u = (Y - B(p,\zeta))/a(p)$, so there is exact aggregation of preferences. Under these conditions, recovery of the market demand functions is sufficient for estimation of WTP, provided all consumers are weighted equally under the social welfare criterion. Thus, the Gorman polar

[13]An immediate implication of (15) and (16) is that the Hicksian and market demand functions are singletons almost everywhere in p.

preference field, with the linear parallel Engle curves restriction that a(p,z,ρ) is independent of (z,ρ) and the restriction that the distribution Φ(z|ρ,y,ζ) is independent of y, yields a simple, relatively easy to identify estimate of mean WTP = WTA from market-level data.

4. An Application: Mean Deadweight Loss from Second-Best Market Organization

Consider pricing in a market with an unregulated retail sector that exhibits Cournot behavior and relies on a common network resource that exhibits increasing returns to scale, such as distribution of gas or electricity from a common transmission system, or supply of internet or telephone services by retailers that connect to a common data transmission network. We consider structuring the network entity as a regulated monopoly, or as a not-for-profit corporation managed by the association of retailers. Table 1 describes the cases we consider.

Table 1

Structure	Organization of Network Entity[14]	Form of Regulation
1	Regulated monopoly, no network-retailer transfers	Optimal regulation with lump-sum network-consumer transfers
2	Regulated monopoly, no network-retailer transfers	Ramsey pricing with zero budget constraint
3	Association-managed not-for-profit, no network-retailer transfers	No regulatory oversight, wholesale prices must be non-negative
4	Association-managed not-for-profit, no network-retailer transfers	No "cross-subsidization" rule that wholesale prices must cover marginal cost
5	Association-managed not-for-profit, no network-retailer transfers	Uniform wholesale markup rate for all goods
6	Unregulated monopoly, network-retailer transfers allowed	None

[14]In all structures, the retailers are unregulated, exhibit Cournot behavior, and as an association seek to maximize the profits of incumbent retailers.

Structure 1 is "optimal" regulation that covers network costs using lump sum transfers from consumers to the network entity. In this structure, the regulator uses wholesale prices to reverse inefficient retail price markups, even though the retail sector is not directly regulated and there are no lump sum transfers between the network entity and the retail sector. Structure 2 is Ramsey (1927) regulation, with no lump sum transfers, and with wholesale prices set to recover network entity costs with minimum deadweight loss. The Ramsey regulation will also attempt to offset inefficient retail price markups. The reasons for considering the association-managed network structures 3-6 are that "user-managed" network entities are a relatively common and accepted form of deregulation, with management by relatively unconcentrated users deemed sufficient to protect the interests of downstream economic agents. In structure 3, we assume the self-regulating restrictions that the network entity be operated on a not-for-profit basis, without "subsidies" to retailers through lump sum transfers or negative wholesale prices. Where some regulation of a partially decentralized industry has been considered necessary, as in the case of mixed regulation where some goods are offered competitively and others are not, a common form of regulatory constraint has been a "no cross-subsidization" or "non-predation" rule that no good be priced below its marginal cost.[15] We consider this case in structure 4. Another proposed form of partial regulation is a requirement that all wholesale markup rates be uniform. We consider this in structure 5. Finally, in structure 6, we consider the worst case for consumers, where the association-controlled network is allowed to monopolize the industry. In the analysis to follow, we will demonstrate that each of these alternatives results in a deadweight loss for consumers relative to structure 1, which achieves a first-best allocation. We show using examples that structure 4 generally involves only a modest deadweight loss for consumers in comparison to the Ramsey pricing structure 2, and in some cases this is also true of structure 3. Structures 5 and 6 on the other hand may involve substantial deadweight losses. These results suggest that from the standpoint of consumer welfare, structure 4, and sometimes structure 3, will be acceptable self-regulating alternatives to Ramsey regulation, particularly when the social costs of regulation and the risks of mis-regulation are taken into account. Conversely, structures 5 and 6 may impose an unacceptable burden on consumers. The

[15]The situation here is not the conventional one in which the regulated entity has a monopoly in some goods and faces competition in others. However, an association-managed not-for-profit network entity's incentives with respect to less elastic and more elastic goods parallels the incentives present in the conventional case.

analysis underlying these conclusions uses the constrained optimization methods of Ramsey (1927), Boiteaux (1956), and Dreze (1964), and is facilitated by assuming a representative consumer with utility of Gorman polar form.

Suppose there are j = 1,...,J commodities provided through the network at wholesale prices w_j, and supplied to consumers at retail prices p_j. Assume that the network has fixed costs F_0 and marginal cost n_j for commodity j. Assume that the retail sector has K firms, each with fixed cost F and marginal cost m_j, in addition to the wholesale cost w_j to them, for selling a unit of commodity j. We will initially assume K fixed and F = 0, so that retailers face constant unit costs and retailer shutdown is not an issue. Later, we consider industry structures when retailers have fixed costs F > 0 and the number of retailers is endogenous. Assume there is a representative consumer with indirect utility described by a Gorman polar form,

$$(19) \quad u = y - \sum_{j=1}^{J} A_j \frac{p_j^{1-\varepsilon_j}}{1-\varepsilon_j},$$

where y is average income, and income and prices are normalized so that a numériare good has price one.[16] Consumer demands have the constant elasticity form $X_j = A_j p_j^{-\varepsilon_j}$. Assume the goods are indexed from least to most elastic, $\varepsilon_1 < \varepsilon_2 < ... < \varepsilon_J$. For simplicity exclude ties and the borderline case of unit elasticity. Assume the number of retailers K is sufficiently large so that $K\varepsilon_1 > 1$.

If the retailer behavior is Cournot, then with common marginal costs for all retailers, retail prices satisfy the markup rules

$$(20) \quad p_j = \frac{m_j+w_j}{1-1/K\varepsilon_j},$$

[16]The dual direct utility function is $u = X_0 + \sum_{j=1}^{J} A_j^{1/\varepsilon_j} X_j^{1-1/\varepsilon_j}/(1-1/\varepsilon_j)$, where X_0 is consumption of the numériare good. If consumption of the numériare good is required to be non-negative, then the condition for the adequacy of income is $y > \sum_{j=1}^{J} A_j p_j^{1-\varepsilon_j}$.

and the profit of the retail sector is

$$(21) \quad \pi = \sum_{j=1}^{J} X_j p_j/K\varepsilon_j - KF = \sum_{j=1}^{J} A_j p_j^{1-\varepsilon_j}/K\varepsilon_j - KF$$

The profit of the network entity is

$$(22) \quad \pi_0 = \sum_{j=1}^{J} X_j(w_j - n_j) - F_0 = \sum_{j=1}^{J} X_j[p_j(1-1/K\varepsilon_j) - M_j] - F_0 \ ,$$

where $M_j = m_j + n_j$ is system marginal cost and we have used the retail markup rule to eliminate $m_j + w_j$. Accounting for the distribution of profit income and any lump-sum transfers, and letting y_0 denote autonomous income, the indirect utility of the consumer is

$$(23) \quad u = y_0 + \sum_{j=1}^{J} X_j p_j/K\varepsilon_j. + \sum_{j=1}^{J} X_j(p_j(1-1/K\varepsilon_j) - M_j) - F_0 - KF - \sum_{j=1}^{J} X_j p_j/(1-\varepsilon_j)$$

$$= y_0 + \sum_{j=1}^{J} X_j(p_j - M_j) - F_0 - KF - \sum_{j=1}^{J} X_j p_j/(1-\varepsilon_j)$$

$$= y_0 - \sum_{j=1}^{J} X_j[p_j\varepsilon_j/(1-\varepsilon_j) + M_j] - F_0 - KF = y_0 - F_0 - KF - \sum_{j=1}^{J} A_j p_j^{-\varepsilon_j}[p_j\varepsilon_j/(1-\varepsilon_j) + M_j] .$$

.

First, suppose the network is operated as a regulated monopoly. Optimal regulation with lump-sum transfers sets wholesale prices so that the resulting retail prices equal system marginal costs. This achieves utility $u = y_0 - F_0 - KF - \sum_{j=1}^{J} A_j M_j^{1-\varepsilon_j}/(1-\varepsilon_j)$.[17] Next consider Ramsey pricing, where wholesale prices are set to maximize (23), given the retail markup rule (20) and the budget constraint that (22) is zero. This is achieved at retail markup rates

[17] If $\varepsilon_1 < 1$, then the industry is essential, and will be operated at any fixed cost F_0, provided autonomous income is high enough to give the consumer positive consumption of the numeriare commodity when goods are priced at system marginal costs. If $\varepsilon_1 > 1$, then it is optimal to shut the industry down if and only if $F_0 > \sum_{j=1}^{J} A_j M_j^{1-\varepsilon_j}/(\varepsilon_j - 1)$. Hereafter, we exclude the shutdown case, and assume there is sufficient autonomous income so that the economy is viable.

$$(24) \qquad \frac{p_j}{M_j} = \frac{1}{1 - \frac{\lambda}{K\varepsilon_j}(K+1-1/\varepsilon_j)},$$

implying wholesale markup rates

$$(25) \qquad \frac{w_j}{n_j} = \frac{1 - \frac{1}{K\varepsilon_j} + \frac{m_j}{n_j}\frac{1}{K\varepsilon_j}(\lambda(K+1-1/\varepsilon_j)-1)}{1 - \frac{\lambda}{K\varepsilon_j}(K+1-1/\varepsilon_j)},$$

where λ is a multiplier that is increased from zero until the budget constraint is satisfied; it is bounded by $0 < \lambda < \min\{1, K\varepsilon_1/(K+1-1/\varepsilon_1)\}$.[18] The Ramsey regulator sets wholesale prices not only to meet the budget constraint for the network, but also to offset deadweight losses from the Cournot markups in the retail sector. The retail markup rate p_j/M_j always exceeds one. For λ fixed, the retail markup is maximized at $\varepsilon_j = 2/(K+1)$, and for larger elasticities declines monotonically. The corresponding wholesale markup rates can be less than one for goods with high elasticities, and can be less than zero when the objective of offsetting Cournot retail markups dominates the objective of network fixed cost recovery for the Ramsey regulator.[19] When $K \to +\infty$, the limiting case of a perfectly competitive retail sector, Ramsey pricing reduces to the classical Ramsey-Boiteaux markup rule $p_j/M_j = 1/(1 - (\lambda/\varepsilon_j))$, so that the least elastic commodities are marked up the most. This limiting case gives the wholesale markup rule $w_j/n_j = [1 + (m_j/n_j)(\lambda/\varepsilon_j)]/(1 - \lambda/\varepsilon_j)$, which always exceeds one and is decreasing in ε_j. The utility obtained at the Ramsey solution is

[18]Ramsey pricing is feasible only if the network entity operating as an unregulated monopoly can cover fixed costs. If $\varepsilon_1 < 1$, this will always be the case. If $\varepsilon_1 > 1$, then fixed costs can be covered if and only if $F_0 > \sum_{j=1}^{J} \left[(1-1/\varepsilon_j)(1-1/K\varepsilon_j)\right]^{\varepsilon_j} A_j M_j^{1-\varepsilon_j}/(\varepsilon_j - 1)$. When $\varepsilon_1 > 1$, note in comparison with the case of optimal regulation with lump sum transfers that there is a range of fixed costs where operation of the network is socially desirable, but Ramsey pricing is infeasible. The reason for this is that the Ramsey regulator cannot recover retailer profit to cover network fixed costs.

[19]For example, if $K = 10$ and $\varepsilon_1 = 1/7$, then $w_1/n_1 = (-2+14\lambda)/(5-14\lambda)$. When F_0 is sufficiently small, one has $\lambda < 1/7$ and $w_1 < 0$. A second example is $K = 10$ and $\varepsilon_1 = 2$, where $w_1 < 0$ if $36/21 < \lambda < 40/21$.

$$(26)\quad u = y_0 - \sum_{j=1}^{J} A_j\left(\frac{M_j}{1-(K+1-1/\varepsilon_j))\lambda/K\varepsilon_j}\right)^{1-\varepsilon_j}[1/(1-\varepsilon_j)-(K+1-1/\varepsilon_j))\lambda/K\varepsilon_j] - F_0 - KF$$

Now consider network pricing when the network entity is a corporation controlled by the retailers' association. First, if the retail sector is perfectly competitive, with K = +∞, then this sector attains zero profit under any wholesale pricing scheme, and will find any feasible scheme that meets the network budget constraint, including the Ramsey pricing solution, acceptable. If K is finite, then the association will seek a wholesale pricing scheme that maximizes (21) subject to the network budget constraint. We consider structures 3-5 where, in 3, the network has only the not-for-profit budget constraint and a non-negative wholesale price constraint, in 4 is bound by an additional "no cross-subsidization" constraint that wholesale markup rates not be less than one, and in 5 is instead bound by an additional regulatory constraint that all wholesale markup rates be the same. In a final structure 6, we drop the network budget constraint, and allow the unregulated network entity to make lump-sum payments to retailers.

Consider the case where all demands are elastic. Then, association profits are maximized at a markup rule

$$(27)\quad p_j/M_j = 1/(1-1/\varepsilon_j)(1+\lambda'/\varepsilon_j),$$

where λ' is a multiplier determined to satisfy the budget constraint. The indirect utility associated with this pricing rule is

$$(28)\quad u = y_0. - \sum_{j=1}^{J} A_j\left(\frac{M_j}{(1-1/\varepsilon_j)(1+\lambda'/\varepsilon_j)}\right)^{1-\varepsilon_j}[\varepsilon_j/(1-\varepsilon_j)+(1+(1-1/\varepsilon_j)(1+\lambda'/\varepsilon_j)] - F_0 - KF.$$

Imposing a regulatory constraint that wholesale markups be at least one is not binding in this case, so structures 3 and 4 give the same results. Structure 5 with a uniform wholesale markup rate is solved by determining a common wholesale markup that balances the network budget, with indirect utility (23) calculated at the resulting retail prices. Structure 6 where the network operates as an unregulated monopoly without a budget constraint results in full monopolization of the industry, as wholesale prices can be set so that retail prices determined by the Cournot markups satisfy the monopoly pricing conditions $p_i = M_i/(1-1/\varepsilon_i)$. Again (23) is used to calculate indirect utility at these prices.

Finally, consider the case where some commodities are inelastic. The effect of a one unit increase in w_j on retail sector profit is $X_j(1-\varepsilon_j)/(K\varepsilon_j-1)$; this is positive if and only if $\varepsilon_j < 1$. The effect of a unit increase in w_j on network profit is $X_j(1-\varepsilon_j+\varepsilon_j M_j/(m_j+w_j)$; this is positive if $\varepsilon_j < 1$ or if $\varepsilon_j > 1$ and $M_j/(m_j+w_j) > 1 - 1/\varepsilon_j$. Lowering the wholesale prices of all commodities j > 1 as far as possible increases retailer profit from the Cournot markup on large sales, and produces large uncovered costs for the network. Then raising the wholesale price of good 1 until the network breaks even generates additional retailer profit. When wholesale prices have lower limits $\underline{w}_j > -m_j$, this yields a corner solution. The indirect utility associated with this pricing rule is

$$(29) \quad u = = y_0 - \frac{A_1}{1-\varepsilon_1}\left[\frac{(m_1+w_1)}{1-1/K\varepsilon_1}\right]^{1-\varepsilon_1} - \sum_{j=2}^{J} \frac{A_j}{1-\varepsilon_j}\left[\frac{(m_j+\underline{w}_j)}{1-1/K\varepsilon_j}\right]^{1-\varepsilon_j} - F_0 - KF.$$

with w_1 set to satisfy

$$(30) \quad F_0 - \sum_{j=2}^{J} A_j\left(\frac{m_j}{1-1/K\varepsilon_j}\right)^{-\varepsilon_j}(\underline{w}_j-n_j) = A_1\left(\frac{m_1+w_1}{1-1/K\varepsilon_1}\right)^{-\varepsilon_1}(w_1-n_1)$$

The results for structures 3 with non-negative wholesale prices and structure 4 with no "cross-subsidization" are obtained, respectively, by fixing $w_j = 0$ and $w_j = n_j$ for j > 1. The results for structure 5 are calculated just as in the case of elastic demands. Structure 6 cannot be analyzed in the case of an inelastic demand, as the network can increase industry profits indefinitely by raising the wholesale price of the inelastically demanded good. This is inconsistent with the domain restrictions on the Gorman form.

The welfare impact of alternative forms of organization for the network entity can be easily calculated from the difference in the indirect utilities, which in the absence of income effects equals WTP = WTA. Two numerical examples with J = 2 illustrate the WTP calculation. Table 2 gives the parametric assumptions – in the first, both goods are elastically demanded, and in the second, one is inelastically demanded. In these examples, the retail sector is relatively concentrated, with K = 10 firms.

Table 2					
Parameter	**Example 1**	**Example 2**	**Parameter**	**Example 1**	**Example 2**
ε_1	1.2	0.2	M_1	2	2
ε_2	5	2	M_2	2	2
n_1	1	1	A_1	23	11
n_2	1	1	A_2	320	40
m_1	1	1	y	1000	1000
m_2	1	1	F_0	10	10

Tables 3 and 4 describe the market outcomes under each network structure, and the WTP for optimal regulation compared with the structure under consideration. In these tables, demands are expressed relative to their levels under optimal regulation (x_i*), the column π/R gives industry profit relative to revenue[20], R/y gives the ratio of industry revenue to income, and WTP/u* gives willingness-to-pay for optimal regulation rather than the structure under consideration, expressed as a ratio to money-metric utility under optimal regulation.

Table 3. Example 1: Elastic Demands, F = 0, Fixed K										
Struct.	K	p_1/M_1	p_2/M_2	w_1/n_1	w_2/n_2	x_1/x_1*	x_2/x_2*	π/R	R/y	WTP/u*
1	10	1.00	1.00	0.83	0.96	1.00	1.00	-25.0%	4.00%	NA
2	10	2.51	1.18	3.60	1.31	0.33	0.44	3.98%	2.60%	0.70%
3	10	3.28	1.04	5.01	1.04	0.24	0.81	4.14%	3.28%	0,94%
4	10	3.28	1.04	5.01	1.04	0.24	0.81	4.14%	3.28%	0,94%
5	10	3.59	3.35	5.58	5.58	0.23	0,00	3.24%	1.57%	1.48%
6	10	6.00	1.25	10.0	1.45	0.12	0.33	8.24%	2.22%	1.81%

[20]For structures 1 and 6, π/R is the ratio of total industry net profit to industry revenue.

Note first that optimal regulation achieves first-best marginal cost pricing at the retail level by setting wholesale markup rates below one, and requires a substantial lump sum transfer from consumers to the network to cover its costs. Ramsey pricing requires substantial wholesale markups, particularly on the less elastic good, and substantially reduces industry consumption. The deadweight loss, expressed as a percentage of optimal money-metric utility, is nevertheless relatively modest. Structure 3 with industry self-regulation leads to a higher wholesale markup on the less elastic good relative to Ramsey pricing, a modestly higher retail profit rate than under Ramsey regulation, and a small additional deadweight loss relative to the Ramsey case. Structure 4 gives the same results as structure 3. On the other hand, structure 5 with a uniform wholesale markup leads to a substantial deadweight loss compared with structure 4. Finally, the unregulated monopoly in structure 6 substantially increases industry profit, and results in an additional deadweight loss.

Table 4. Example 2: One Demand Inelastic, F = 0, Fixed K

Struct.	K	p_1/M_1	p_2/M_2	w_1/n_1	w_2/n_2	x_1/x_1^*	x_2/x_2^*	π/R	R/y	WTP/u*
1	10	1.00	1.00	0.00	0.90	1.00	1.00	-25.5%	3.92%	NA
2	10	3.15	1.14	2.15	1.16	0.79	0.78	63.5%	6.56%	0.36%
3	10	9.56	0.53	8.56	0.00	0.64	3.61	153%	15.5%	3.41%
4	10	3.33	1.05	2.33	1.00	0.79	0.90	66.4%	6.91%	0.37%
5	10	2.81	1.46	1.81	1.81	0.81	0.46	57.6%	5.73%	0.47%

In example 2, Ramsey pricing loads most of the recovery of network costs on the inelastic good, achieving the budget constraint with a retail price for the elastic good that is close to its system marginal cost. Structure 4 with the constraint that wholesale prices meet network marginal costs achieves nearly the efficiency of Ramsey pricing. Structure 5 with a uniform markup requirement entails an additional deadweight loss. However, structure 3 in which the association manages the network entity with only a non-negative wholesale price constraint is able to substantially increase retail sector profit even with the not-for-profit budget constraint, by setting a wholesale price of zero for the elastic good, allowing it to be sold at high volume and a small retail markup. This wholesale "loss-leader" then allows a high wholesale markup on the inelastic good while meeting the budget constraint, and this is translated into a profitable retail markup on this good.

Up to now, we have assumed retail firms had zero fixed cost so that shut-down was not an issue, and the number K of these firms was fixed. Now assume instead that retail firms do have positive fixed costs, and K is determined endogenously by entry and exit. Economic questions are the incentives of an association of incumbent retailers in the management of a not-for-profit common resource, and what a Ramsey regulator assumes about retail entry and its effect on welfare. An initial observation is that the association of incumbents will, if possible, raise barriers to entry, allowing its members to earn some economic rents. If it can, it will do this by restricting entrant's access to the common network resource or making "buy-in" costly, by limiting entrant's management rights in network operations, and by increasing wholesale prices to entrants relative to those for incumbents. We will assume here that incumbents cannot restrict entry by any of these methods, and that K is the largest integer at which retailer profit $\pi = \sum_{j=1}^{J} A_j p_j^{1-\varepsilon_j}/K\varepsilon_j$ - KF, where F is fixed cost per retailer, is non-negative. We will also assume that the retailers' association and the regulator treat K as fixed in their optimizing calculations, a Nash assumption on play in the entry/exit game. Note that optimal or Ramsey regulation then fails to account for the entry/exit margin, and the social incentives to reduce K to lower retail sector fixed costs and to increase K to reduce Cournot profit margins. Consequently, structure 1 with optimal regulation is no longer first-best when K > 1. We recalculate examples 1 and 2 with the assumption that each retailer has a fixed cost F such that K = 10 in structure 1, The results are given in Tables 5 and 6.

Table 5. Example 1: Elastic Demands, F = 0.2, K endogenous

Struct.	K	p_1/M_1	p_2/M_2	w_1/n_1	w_2/n_2	x_1/x_1^*	x_2/x_2^*	π/R	R/y	WTP/u*
1	10	1.00	1.00	0.83	0.96	1.00	1.00	-30.0%	4.00%	NA
2	8	2.71	1.20	3.66	1.33	0.30	0.41	0.88%	2.62%	0.76%
3	8	3.45	1.06	5.19	1.07	0.23	0.74	1.05%	3.13%	0.97%
4	8	3.45	1.06	5.19	1.07	0.23	0.74	1.05%	3.13%	0.97%
5	7	5.77	5.23	9.16	9.16	0.12	0.00	0.70%	1.41%	2.04%
6	8	6.00	1.25	9.75	1.44	0.12	0.33	4.24%	2.22%	1.78%

Table 6. Example 2: One Demand Inelastic, F = 1.0, K endogenous										
Struct.	K	p_1/M_1	p_2/M_2	w_1/n_1	w_2/n_2	x_1/x_1^*	x_2/x_2^*	π/R	R/y	WTP/u*
1	10	1.00	1.00	0.00	0.90	1.00	1.00	-51.1%	3.92%	NA
2	14	2.46	1.10	2.17	1.08	0.84	0.83	2.49%	5.76%	0.62%
3	20	6.14	0.51	8.21	0.00	0.70	3.80	3.64%	12.1%	3.88%
4	14	2.53	1.04	2.26	1.00	0.83	0.93	2.76%	5.96%	0.62%
5	22	2.81	1.48	1.81	1.81	0.81	0.46	1.41%	5.73%	1.71%

Note that for structure 1, π/R is the ratio of total industry profits, including the network, to industry revenue; the profits of the retail sector are non-negative and the regulator makes no lump sum transfers to this sector. When K is endogenously determined by entry and exit, industry self-regulation tends to decrease K when demands are elastic, and increase K when a demand is inelastic. However, the welfare consequences of alternative structures are very similar, with structure 4 nearly as good as Ramsey pricing in all circumstances, and structures 3 and 5 substantially worse in some circumstances.

These examples serve to illustrate the usefulness of the Gorman representative consumer assumption for welfare analysis, and estimation of the benefits (which must be weighed against the costs) of various levels of intervention in industrial structure, e.g., full Ramsey regulation of a common network entity, or a legal requirement that wholesale prices at least cover network marginal costs. From these examples, we conclude that it is possible to achieve a reasonable welfare outcome under industry self-regulation of a common network resource. If the goods sold by the industry have elastic demands, then network operation as a non-for-profit enterprise with non-negative wholesale prices is sufficient without regulatory supervision. However, if some of the goods have inelastic demands, then it is possible for the retailer's association to manipulate wholesale prices within the network's budget constraint to substantially increase retail profits, with a substantial deadweight loss. This can be avoided by imposing a "no cross-subsidization" regulatory constraint that all wholesale prices cover marginal costs. An alternative form of supervision, requiring uniform wholesale markup rates, generally entails larger deadweight losses for consumers.

5. Consumers in Space

Welfare analysis using the Gorman polar preference field can handle heterogeneity in tastes, income, and non-market environments. It does so most easily when Engle curves are affine linear and parallel for different ρ and z, and the distribution of z is exogenous and hence independent of income, so that market demands are consistent with the preferences of a representative consumer. Even without the representative consumer characterization, when individual Gorman polar demand functions can be recovered and are sufficient to identify all the effects of non-market goods on utility, it remains easy to compute individual WTP, and from this deduce the distribution of WTP in the population, or moments or quantiles of this distribution. With the reinterpretation that follows, this preference field also facilitates analysis of consumer choice in space.

Consumers may face location or address choices in hedonic or physical space, and may be heterogeneous in their endowed tastes or initial location. The indirect utility of location t in a choice set T may be written in general as

$$(31) \quad u = V(p,y-r(t),z(t),t,\rho),$$

where r(t) is the cost of location t and z(t) describes the non-market environment at this location.[21] The function (31) can be interpreted as the maximum utility obtainable in the market, conditioned on location t. This function inverts to an expenditure function

$$(32) \quad y - r(t) = M(p,u,z(t),\rho).$$

The consumer will choose a location to maximize (31),

$$(33) \quad u = V^*(p,\underline{r},y,\underline{z},\rho) \equiv \max_{t\in T} V(p,y-r(t),z(t),t,\rho),$$

[21] Market prices p are assumed independent of location in this formulation, but this is not essential, and can be generalized even within the current notation by letting the location determine the subset of market goods that are feasible to consume.

where $\underline{r}$, $\underline{z}$ denote the cost and environment functions on T.[22] The associated expenditure function is

(34) $y = M^*(p,\underline{r},u,\underline{z},\rho) \equiv \min_{t\in T} \{r(t) + M(p,u,z(t),t,\rho)\}$.

Combining (33) and (34), a money-metric utility function for the consumer in space is

(35) $u = \mu^*(p',\underline{r}',\underline{z}';p,\underline{r},y,\underline{z},\rho) = \min_{t\in T} \max_{s\in T} \mu_{st}(p',\underline{r}',\underline{z}';p,\underline{r},y,\underline{z},\rho)$

where

(36) $\mu_{st}(p',\underline{r}',\underline{z}';p,\underline{r},y,\underline{z},\rho) = r(t) + M(p',V(p,y\text{-}r(s),z(s),s,\rho),z'(t),t,\rho)$

is the s-location, t-metric money-metric utility.

Equation (32), or equivalently, (36), with ρ interpreted as a random effect that varies across (or within) members of the population, is a *random utility maximization* (RUM) model; see McFadden (1973, 1981). It is simplest to think of T as a finite set, giving a discrete choice problem, but in general we will require only that T be a compact subset of a finite-dimensional address space, that $\underline{z}$ be contained in a compact space of functions from T into Z, and that $\underline{r}$ be a point in the Banach space $B(T,\mathcal{T})$ of uniform limits of real-valued linear combinations of characteristic functions on the Borel σ-field $\mathcal{T}$ of T. Let $\delta(\cdot;p,\underline{r},y,\underline{z},\rho)$ be a probability measure on $\mathcal{T}$, defined when the set of maximands in (33) is non-empty, whose support is contained in the set of maximands; it is an indicator for choice when there is a unique maximand, and is otherwise interpreted as the random device used by the consumer to break ties. Note that Roy's identity holds when there is a unique maximand in t, so that

(37) $\delta(t;p,\underline{r},y,\underline{z},\rho) = - \nabla_{r(t)}V^*(p,\underline{r},y,\underline{z},\rho) / \nabla_y V^*(p,\underline{r},y,\underline{z},\rho)$.

---

[22] The properties of indirect utility functions (quasi-convex and homogeneous of degree zero in $(p,\underline{r},y)$, increasing in y, non-increasing in $p,\underline{r}$) are preserved by the operation of maximization over a set T.

Now consider a population of consumers with heterogeneous tastes ρ composed of components ε(t) that vary with t, and components η that do not, so that $\rho = (\eta,\underline{\varepsilon})$. Suppose income y and the non-market profile $\underline{z}$ are statistically independent of $\underline{\varepsilon}$, given η This assumption will hold, for example, if η is a sufficient statistic for the mechanism that assigns y and $\underline{z}$. Then choice probabilities in the population, conditioned on η, y, and $\underline{z}$, satisfy

$$(38)\quad P_T(A|p,\underline{r},y,\underline{z},\eta) = \int_{\underline{\varepsilon}} \delta(A;p,\underline{r},y,\underline{z},\eta,\underline{\varepsilon})\Psi_\varepsilon(d\underline{\varepsilon}|\eta)$$

for $A \in \mathrm{T}$. The integral in (38) can be non-trivial to calculate even when T is finite.[23] Thus, it is useful to consider cases where (38) has a tractable form. A useful simplification is to consider conditional indirect utility functions of the additively separable form

$$(39)\quad u = V(p,y\text{-}r(t),z(t),t,\eta,\varepsilon(t)) = V_1(p,y\text{-}r(t),z(t),t,\eta) + \varepsilon(t),$$

where V_1 is independent of ε(t) and ε has a distribution $\Psi_\varepsilon(\underline{\varepsilon}|\eta)$ that gives a tractable integral in (38).[24] An alternative simplification that is very useful is to assume that preferences have a Gorman polar form,

$$(40)\quad u = V(p,y\text{-}r(t),z(t),t,\eta,\varepsilon(t)) = [y - r(t) - b(p,z(t),t,\eta,\varepsilon(t))]/a(p,z(t),t,\eta),$$

where a(p,z(t),t,η) and b(p,z(t),t,η,ε(t)) are linear homogeneous and concave in p. A further assumption, that the function a(p,z(t),t,η) is independent of any variables that vary with location, a(p,η); induces a "parallel Engle curves" property across locations. The form (40) with a(p,η) independent of location is also termed an *Additive Income Random Utility Maximization* (AIRUM) model. This leads to a *locationally representative* consumer whose demands at various

---

[23]Technical machinery is needed to define (38) when T is infinite, but intuitively it can be characterized as the probability limit of an average of the integrand over trajectories $\underline{\varepsilon}$ drawn by simulation from the distribution $\Psi_\varepsilon(\underline{\varepsilon}|\eta)$.

[24]Defining $\varepsilon^*(t) = V(p,y\text{-}r(t),z(t),t,\eta,\varepsilon(t)) - V(p,y\text{-}r(t),z(t),t,\eta,0)$ and renaming ε*(t) as ε(t) gives the form (40) without loss of generality, but the requirement that this disturbance have a specified distribution $\Psi_\varepsilon(\underline{\varepsilon}|\eta)$ independent of the other variables in the problem is a substantive restriction on the preference field.

locations are the choice probabilities. Note that in general we do not require in (40) that the taste disturbance $\varepsilon(t)$ be additive and linear. There is a common model nested within (39) and (40): Suppose $a(p,z(t),t,\eta)$ in (40) is linear in p, and the function b has the additively separable form $b(p,z(t),t,\eta,\varepsilon(t)) = b_1(p,z(t),t,\eta) + a(p,z(t),t,\eta)\varepsilon(t)$. Then indirect utility satisfies

(41) $u = V(p,y\text{-}r(t),z(t),t,\eta,\varepsilon(t)) = [y - r(t) - b_1(p,z(t),t,\eta)]/a(p,z(t),t,\eta) + \varepsilon(t).$

We will first analyze spatial demand models in what we will term the Generalized Extreme Value case, based on (39) and what is called the GEV family of distributions for the disturbances. After this, we analyze what we will term the locationally representative Gorman consumer case, based on(40) and a locationally parallel Engle curves restriction.

Generalized Extreme Value Case

Consider (39) with $T = \{t_1,...,t_J\}$ finite, and assume that the additively separable taste disturbance $\varepsilon(t)$ has a Generalized Extreme Value (GEV) distribution. To define this distribution, first define a *GEV generating function* $H(w_1,...,w_J|\eta)$: a non-negative linear homogeneous function of non-negative $(w_1,...,w_J)$ with mixed partials that satisfy an alternating sign condition $(-1)^j\partial^j H/\partial w_1...\partial w_j \le 0$ for j = 1,...,J. Term such a function S-*proper* for a subset S of T if $H(\mathbf{1}_j|\eta) > 0$ for all $j \in S$, and $H(\mathbf{1}_{T\backslash S}|\eta) = 0$. A GEV distribution is of the form

(42) $\Psi_\varepsilon(\underline{\varepsilon}|\eta) = \exp(-H(\exp(-\varepsilon_1),...,\exp(-\varepsilon_J)|\eta),$

where H is a T-proper GEV generating function.[25] As a notational shorthand, let $u_j = u(t_j)$, $v_j = V_1(p,y\text{-}r(t_j),z(t_j),t_j,\eta)$, and $\varepsilon_j = \varepsilon(t_j)$. Let $\gamma = 0.5772$ denote Euler's constant, and Γ denote the gamma function. A standardized univariate Extreme Value Type 1 (EV1) distribution function has the form $\exp(-\exp(-\varepsilon))$. The following result, from McFadden (1978) and Beirlaire, Bolduc, and McFadden (2003), shows that GEV distributions yield closed-form choice probabilities:

[25] This specification is intrinsically finite. However, it is possible, using the methods of Dagsvik (1994,1995), to construct extreme value processes on a continuum, and use the properties in Lemma 2 below to build up generalized extreme value processes that generate closed form choice probabilities on a continuum.

Theorem 1. If $H(w_1,...,w_J\ |\eta)$ is a T-proper GEV generating function, then a random vector $(u_1,...,u_J)$ satisfying (39), with $\underline{\varepsilon}$ satisfying (42), has the properties

A. The u_j for j = 1,...,J are EV1 with common variance $\pi^2/6$, means $v_j + \log H(\mathbf{1}_j\ |\eta) + \gamma$, and moment generating functions $\exp(\xi v_j)H(\mathbf{1}_j)^{\xi}\Gamma(1-\xi)$.

B. $u_0 = \max_{i=1,...,J} u_i$ is EV1 with variance $\pi^2/6$, mean $\log H(\exp(v_1),...,\exp(v_J)) + \gamma$, and moment generating function $H(\exp(v_1),...,\exp(v_J))^{\xi}\Gamma(1-\xi)$.

C. Letting $H_j(w_1,...,w_J|\eta) = \partial H(w_1,...,w_J|\eta)/\partial w_j$, the probability P_j that $j = \mathrm{argmax}_{i\in\mathbf{T}}\ u_i$ satisfies

(43) $$P_j = P_T(t_j|p,\underline{r},y,\underline{z},\eta) = \frac{\exp(v_j)\cdot H_j(\exp(v_1),...,\exp(v_J)|\eta)}{H(\exp(v_1),...,\exp(v_J)|\eta)}.$$

The linear function $H(w) = w_1 + ... + w_J$ is a GEV generating function; the vector $(\varepsilon_1,...,\varepsilon_J)$ given by (42) for this H has independent extreme value distributed components. The choice probabilities (43) then have a *multinomial logit* (MNL) form,

(44) $$P_j = \exp(v_j)/\Sigma_{i\in\mathbf{T}} \exp(v_i).$$

The next result gives operations on GEV generating functions that can be applied recursively to generate additional GEV generating functions.

Lemma 2. The family of GEV generating functions is closed under the operations:

A. If $H(w_1,...,w_J\ |\eta)$ is a S-proper GEV generating function, $S \subseteq T$, then $H(\alpha_1 w_1,...,\alpha_J w_J\ |\eta)$ for $\alpha_1,...,\alpha_J \geq 0$ and $B = \{j\in S|\alpha_j > 0\}$ is a B-proper GEV generating function..

B. If $H(w_1,...,w_J\ |\eta)$ is an A-proper GEV generating function and $\sigma > 1$, then $H(w_1^{\sigma},...,w_J^{\sigma}\ |\eta)^{1/\sigma}$ is an A-proper GEV generating function for $A \subseteq T$.

C. If $H^{A}(w_1,...,w_J\,|\eta)$ and $H^{B}(w_1,...,w_J\,|\eta)$ are, respectively, A-proper and B-proper GEV generating functions, where A, B $\subseteq$T are not necessarily disjoint, then $H^{A}(w_1,...,w_J\,|\eta) + H^{B}(w_1,...,w_J\,|\eta)$ is a A$\cup$B-proper GEV generating function.

.

McFadden and Train (2000) show for discrete choice that any regular random utility can be approximated as a η-mixture of GEV-distributed utilities, and thus market-level discrete choice probabilities are approximately η-mixtures of (44). Note that choice probabilities based on Theorem 1 may depend on income, and thus capture choice behavior that varies with income.

Locationally Representative Gorman Consumer Case

A preference restriction leading to tractable choice probabilities is the Gorman polar form (40) with a(p,η) independent of location. The expectation with respect to the distribution $\Psi_\varepsilon(\underline{\varepsilon}|\eta)$ of the maximum of these conditional utility functions over location is a representative consumer utility function that is again of Gorman polar form,

$$(45)\quad u = V^*(p,\underline{r},y,\underline{z},\eta) \equiv \int_{\varepsilon} \max_{t\in T}\,[y - r(t) - b(p,z(t),t,\eta,\varepsilon(t))]\Psi_\varepsilon(\underline{d\varepsilon}|\eta)/a(p,\eta)$$

$$= y/a(p,\eta) + G(p,\underline{r},\underline{z},\eta),$$

where

$$(46)\quad G(p,\underline{r},\underline{z},\eta) = \int_{\varepsilon} \max_{t\in T}\,[\,-\,r(t)/a(p,\eta) - b(p,z(t),t,\eta,\varepsilon(t))]\Psi_\varepsilon(d\underline{\varepsilon}|\eta)/a(p,\eta).$$

The demands obtained from (45) by applying Roy's identity to the price vector $\underline{r}$ are the choice probabilities (38): $P_T(t|p,\underline{r},y,\underline{z},\eta) = -a(p,\eta)\partial G(p,\underline{r},\underline{z},\eta)/\partial r(t)$. The function $G(p,\underline{r},\underline{z},\eta)$ in (46) is termed a *social surplus* function; it necessarily has the properties

(i) G is homogeneous of degree zero, convex, and non-increasing in $(p,\underline{r})$,

(ii) For any scalar θ, $G(p,\underline{r}+\theta,\underline{z},\eta) = G(p,\underline{r},\underline{z},\eta) - \theta$,

(iii) All the mixed partials of G with respect to $\underline{r}$ exist and are non-positive.

These properties are also sufficient for a function G to satisfy the construction (46) for some distribution $\Psi_{\varepsilon}(\underline{\varepsilon}|\eta)$. The essential elements of this characterization are due to Williams (1977) and Daly and Zachery (1978); the result is proved in McFadden (1981), who notes its close relation to the aggregation properties of the Gorman polar form:

Theorem 3 [Williams-Daly-Zachery]. Consider the field of additive-income random utility models (AIRUM) of the form (40) with $a(p,z(t),t,\eta)$ and $b(p,z(t),t,\eta,\varepsilon(t))$ positive, non-decreasing, linear homogeneous, and concave in p, The mapping (45) from AIRUM preferences is onto the class of social surplus functions satisfying (i)-(iii).

In the case of the preference field (41) that is common to both the Generalized Extreme Value and the Locationally Representative Gorman Consumer cases, both Theorem 1 and Theorem 2 apply, and establish that

(47) $G(p,\underline{r},\underline{z},\eta) = \log H(\exp(-[r_1\text{-}b_1(p,z_1,t_1,\eta]/a(p,\eta)),...,\exp(-[r_J\text{-}b_1(p,z_J,t_J,\eta]/a(p,\eta))) + \gamma,$

with choice probabilities satisfying (43) with $v_j = -[r_j\text{-}b_j(p,z_j,t_j,\eta]/a(p,\eta)$. In this combined case, the choice probabilities are independent of income. When this case is realistic for applications, it facilitates welfare analysis by eliminating income effects.

6. Consumer Welfare at the Extensive Margin

The welfare analysis problem for the consumer in space can be stated as that of determining mean WTP or WTA, or quantiles of the distributions of WTP or WTA, for a change from $(p',\underline{r}',y',\underline{z}')$ to $(p'',\underline{r}'',y'',\underline{z}'')$, based on individual or market-level observations on choice behavior. McFadden (1999) analyzes the problem of estimating or bounding WTP and WTA for this problem; the following summary is most easily interpreted when T is finite, but finiteness is not essential. For any $s,t \in T$, define C_{st} as the net compensating variation (i.e., net reduction in final income) that makes a consumer indifferent to location s before the change and location t after the change; i.e.,

$$(48) \quad C_{st} = \mu_{st}(p'',\underline{r}'',\underline{z}'';p'',\underline{r}'',y'',\underline{z}'',\rho) - \mu_{st}(p'',\underline{r}'',\underline{z}'';p',\underline{r}',y',\underline{z}',\rho)$$
$$= y'' - r''(t) - M(p'',V(p',y'-r'(s),z'(s),s,\rho),z''(t),t,\eta,\underline{\varepsilon}).$$

Define C* to be the compensating variation that equates *maximum* utility over T before and after the change. Then

$$(49) \quad C^* = y'' - \mu^*(p'',\underline{r}'',\underline{z}'';p',\underline{r}',y',\underline{z}',\rho) = y'' - \min_{t\in T} \max_{s\in T} \mu_{st}(p',\underline{r}',\underline{z}';p,\underline{r},y,\underline{z},\rho)$$
$$= \max_{t\in T} \min_{s\in T} C_{st}.$$

Hence, the optimal locations s′ before the change and t″ after the change satisfy

$$(50) \quad C_{s's'} \le C^* \le C_{t''t''}.$$

Define B_{st} to be the event that s is the optimal location before the change and t is the optimal location after the change, and define $B_{s\cdot} = \cup_{t\in T} B_{st}$ and $B_{\cdot t} = \cup_{s\in T} B_{st}$. Let $P'(s)$ denote the uncompensated location choice probability before the change, and $P''^*(t)$ denote the choice probability after the change *when compensation* C* *is paid*, from (50). In general, the quantities $C_{s's'}$, C*, and $C_{t''t''}$ all depend on $\underline{\varepsilon}$. Mean WTP, conditioned on (η, $p',y',\underline{r}',\underline{z}',p'',\underline{r}'',y'',\underline{z}''$), satisfies the bounds

$$(51) \quad \int_s \mathbf{E}(C_{ss}|B_{s\cdot})P'(ds) \le WTP \le \int_s \mathbf{E}(C_{ss}|B_{\cdot s})P''^*(ds),$$

where $\mathbf{E}(C_{ss}|B_{s\cdot})$ denotes conditional expectation given the event $B_{s\cdot}$. A completely analogous development in terms of constant-location equivalent variations E_{ss} yields similar inequalities,

$$(52) \quad \int_s \mathbf{E}(E_{ss}|B_{s\cdot})P'^*(ds) \le WTA \le \int_s \mathbf{E}(E_{ss}|B_{\cdot s})P''(ds),$$

where P'' denotes the uncompensated location choice probability after the change, P'^* denotes the choice probability before the change when compensation equal to the overall equivalent variation E* is given, and the events B_{st} and their unions are now defined with equivalent rather than compensating adjustments.

A substantial simplification of (51) and (52) occurs when preferences have the specification (39), with additively separable taste disturbance ε(t), as in this case the constant-location compensating and equivalent variations do *not* depend on ε and one has $\mathbf{E}(C_{ss}|B_{s\cdot}) = C_{ss}$ and $\mathbf{E}(E_{ss}|B_{\cdot s}) = E_{ss}$. Since the left-hand inequality in (51) and the right-hand inequality in (52) do not require knowing the overall compensating and equivalent variations C* and E*, these bounds are relatively straightforward to compute. Then, (51) and (52) may provide easy bounds that are sufficiently tight to guide policy without recovering individual WTP and WTA. Note that the terms in these inequalities can be integrated with respect to the distributions $\Phi(\underline{z}|\eta,y,\zeta)\Gamma(y|\eta,Y)\Psi_\eta(\eta)$ to obtain bounds on the overall population mean WTP and WTA. Alternately, lower and upper bounds on the distribution of WTP and WTA can be obtained by noting that

$$(53)\quad \int_s \mathbf{1}(C_{ss} \ge \alpha)P'(ds) \le \text{Prob}(\ \alpha \le \text{WTP} \le \beta) \le \int_s \mathbf{1}(C_{ss} \le \beta)P''^*(ds)$$

and

$$(54)\quad \int_s \mathbf{1}(E_{ss} \ge \alpha)P'^*(ds) \le \text{Prob}(\ \alpha \le \text{WTA} \le \beta) \le \int_s \mathbf{1}(E_{ss} \le \beta)P''(ds).$$

Thus, a project is desirable by a mean benefit-cost criterion if $0 \le \int_s C_{ss}P'(ds)$, and undesirable if $0 \ge \int_s E_{ss}P''(ds)$, while this project is desirable by a median voter criterion if , $0.5 \le \int_s \mathbf{1}(C_{ss} \ge 0)P'(ds)$ and undesirable if $0.5 \ge \int_s \mathbf{1}(E_{ss} \le \beta)P''(ds)$.

When T is finite and preferences have the form (39) with generalized extreme value distributed additive taste disturbances, Theorem 1 gives closed forms for the choice probabilities in (51) and (52), and also implies a "certainty equivalent" compensating variation $C^{\#}$ satisfying

$$(55)\quad H(e^{v_1''^*},...,e^{v_1''^*}|\eta) = H(e^{v_1'},...,e^{v_1'}|\eta),$$

where $v_j' = V_1(p',y'-r'(t_j),z'(t_j),t_j,\eta)$ and $v_j''^* = V_1(p'',y''-r''(t_j)-C^{\#},z''(t_j),t_j,\eta)$. The bounds (50) hold for $C^{\#}$ as well as for C*, but in general these quantities are not equal, and $C^{\#}$ is a biased estimate of C*. The exception where $C^{\#}$ equals C* occurs when V_1 is linear in y, a case that corresponds to the preference field (40).

Next consider preferences that have the Gorman polar form (41). Then the constant-location compensating and equivalent variations have explicit forms,

$$(56)\quad C_{ss} = a(p'',z''(s),s,\eta)\left[\frac{y''-r''(s)-b(p'',z''(s),s,\eta)}{a(p'',z''(s),s,\eta)} - \frac{y'-r'(s)-b(p',z'(s),s,\eta)}{a(p',z'(s),s,\eta)}\right],$$

$$(57)\quad E_{ss} = a(p',z'(s),s,\eta)\left[\frac{y''-r''(s)-b(p'',z''(s),s,\eta)}{a(p'',z''(s),s,\eta)} - \frac{y'-r'(s)-b(p',z'(s),s,\eta)}{a(p',z'(s),s,\eta)}\right].$$

These provide relatively convenient bounds in (51) or (52). However, the greatest simplification comes when $a(p,z(s),s,\eta)$ is independent of location. Then, location minimizes $[r(t) + b(p,z(t),t,\eta)]/a(p,\eta) - \varepsilon(t)$, independently of y, and the exact compensating variation has the explicit form

$$(56)\quad C^* = a(p'',\eta)\left\{\begin{array}{l}\left(\dfrac{y''}{a(p'',\eta)} - \dfrac{y'}{a(p',\eta)}\right) - \min_{s\in T}\left(\dfrac{r''(s)+b(p'',z''(s),s,\eta)}{a(p'',\eta)} - \varepsilon(s)\right) \\ \qquad + \min_{s\in T}\left(\dfrac{r'(s)+b(p',z'(s),s,\eta)}{a(p',\eta)} - \varepsilon(s)\right)\end{array}\right\}.$$

When the set T is finite, and the vector of additive taste disturbances $\underline{\varepsilon}$ has a GEV distribution with GEV generating function H, then from Theorem 1,

$$(59)\quad \text{E C}^* = a(p'',\eta)\log\left\{\frac{H(e^{v_1''},\ldots,e^{v_J''}|\eta)}{H(e^{v_1'},\ldots,e^{v_J'}|\eta)}\right\},$$

where $v_j'' = [y'' - r''(t_j) - b(p'',z''(t_j),t_j,\eta)]/a(p'',\eta)$ and $v_j' = [y' - r'(t_j) - b(p',z'(t_j),t_j,\eta)]/a(p',\eta)$. This formula can be mixed with respect to the distribution $\Gamma(y|\eta,Y)\Psi_\eta(\eta)$ to give mean compensating variation for the population, and a project is desirable by this criterion if and only if

$$(60)\quad 0 < \int_\eta \log\left\{\frac{H(e^{v_1''},\ldots,e^{v_J''}|\eta)}{H(e^{v_1'},\ldots,e^{v_J'}|\eta)}\right\}\Gamma(dy|\eta,Y)\Psi\eta(d\eta).$$

This is the same criterion that one would obtain from a $\Gamma(y|\eta,Y)\Psi_\eta(\eta)$ population mixture of locationally representative Gorman polar consumers with indirect utility functions

$$(62)\quad u = \frac{y}{a(p|\eta)} + \log H(e^{v_1^*},\ldots,e^{v_J^*}|\eta),$$

where $v_j^* =[- r(t_j) - b(p,z(t_j),t_j,\eta)]/a(p,\eta)$.

7. An Application: Consumer Harm from Proximity to a Hazardous Waste Site

Suppose consumers choose residential location to maximize utility, taking into account the price and hedonic attributes of houses in the market. Heterogeneity in consumer tastes will lead them to sort themselves out over locations, with the prices at various locations adjusting to clear the market. Suppose it is revealed that a neighborhood is in the proximity of a hazardous waste site that has produced groundwater contamination.

A new equilibrium will be reached, with housing prices adjusting and consumers relocating in response to concerns about the contamination. The economic questions are WTP or WTA for remediation of the contamination, and the conditions under which these welfare measures can be identified from market observations. The analysis that follows could also be applied to determine the welfare effects of an urban transportation system improvement, such as the addition of a link to the transportation network, that has spatially distributed effects. The translation of language for this application is left to the reader.

Current practice in natural resource damage assessment is to use one of three measurement approaches:

(1) The hedonic price method (HPM) in which housing prices are regressed on dwelling, neighborhood, and overall market attributes, and variables that capture permanent neighborhood effects, background time effects, and interaction effects that are intended to capture the impact of the contamination announcement on the affected neighborhood.

(2) The travel cost method (TCM) in which location decisions of individual consumers are estimated as discrete choice functions of housing and neighborhood attributes, dwelling prices, and indicators for neighborhood, time, and interaction effects.

(3) The stated preference method (SPM) in which WTP for changes in environmental conditions is elicited in experiments that offer hypothetical market choices that include environmental goods that in reality do not have markets.

Initial questions are how these measures are related to WTP and WTA, and what theoretical restrictions and data circumstances will allow us to estimate or bound the impact of the revealed contamination on consumer welfare.

I will not be concerned in this paper with SPM and TCM. Clearly, SPM experimental methods can recover all aspects of preferences, provided the hypothetical setting and incentives can be structured to elicit responses consistent with preferences and behavior in the real world. The challenge is to define such experiments. The TCM can recover all features of utility that vary with location. If for example consumers have preferences of the Gorman polar form (41) with additive GEV taste disturbances $\varepsilon(t_j)$, then exact expected compensating variation is given by (59), which depends on the expression $v_j = [y - r(t_j) - b(p,z(t_j),t_j,\eta)]/a(p,\eta)$ evaluated before and after a change. Components of this formula that do not vary with t, but do vary with the change, cannot be identified by the TCM, but may be identified by this method augmented with estimates of the market demand for the commodities with price vector p. Typically in applications, the environmental change is assumed to leave invariant variables that do not vary with t, so that identification of expected WTP is achieved by TCM. Separability assumptions on the preference field may be important in achieving identification in the presence of effects that do not depend on t.

I turn now to the hedonic price method. The basic idea of this method is that if two properties, one in a "treatment" area affected by the environmental hazard, the other in an unaffected "control" area, are *identical in all other respects*, then the difference in their market prices reflects the value consumers place on the hazard. The controls might include properties in a nearby neighborhood that are not exposed to the hazard, or properties in the affected neighborhood at a time when the hazard is absent, or both. The "value" measured by the HPM is a *pecuniary* loss, which is not in general the same as the monetized loss in utility, or WTP, of residents in the affected neighborhood. We ask how the two differ, and when the second can be recovered from information on the first.

If properties and neighborhoods differ in their attributes over location or time, then these differences confound observation of the effect of the hazard. However, it is possible to control and correct for the effect of *observed* attributes by use of an appropriately specified *hedonic regression*. As a practical matter, successful isolation of the effects of the hazard from possible confounders requires accurate measurement of attributes, search for the appropriate functional form and the variable transformations and interactions it requires, and careful attention to experimental design to identify and isolate significant confounders. To control possible confoundment from both persistent neighborhood effects and time variation in the overall market, current "best practice" in environmental economics is to use a two-way experimental design, with observations in control neighborhoods unaffected by the hazard, as well as the affected neighborhood, both during the period when the hazard is present and before it appeared. These observations are analyzed in a regression model

(63) $$y_{wnt} = \alpha_n + \gamma_t + \delta_w + x_{wnt}\beta + \varepsilon_{wnt},$$

where

n neighborhood (n = 1 for treated, n = 0 for controls)

t	time (t = 0 before hazard, t = 1 during hazard)
w	hazard indicator (w = n·t)
y	log sales price
x	measured property and neighborhood attributes
ϵ	disturbance

In this model, α_n is a persistent neighborhood effect, γ_t is an overall time effect, and δ_w is interpreted as the average percent diminution in values of the affected properties attributable to the hazard. The reason for introducing the apparently superfluous hazard indicator w = n·t is that later we will consider the counterfactual $w \equiv 0$, and the counterfactual values x_{011} and y_{011}. Normalizations $\alpha_0 = \gamma_0 = \delta_0 = 0$ are imposed for identification.

In the terminology of the analysis of treatment effects, the mean effect of the hazard on properties in the affected neighborhood is called the *average effect of treatment on the treated*, given x, and is defined as

(64) $\mathbf{E}\{TT|x\} = \mathbf{E}\{y_{111}|x\} - \mathbf{E}\{y_{011}|x\}$,

where y_{111} is log sales price in the factual case that treatment (exposure to the hazard, or w = 1) occurs in the affected neighborhood n = 1 at time t = 1, and y_{011} is the (unobserved) log sales price in the counterfactual case that there is no hazard (w = 0) at n = t = 1. Under some assumptions, the counterfactual conditional expectation can be expressed as a linear combination of conditional expectations that can be estimated from observations,

(65) $\mathbf{E}\{y_{011}|x\} = \mathbf{E}\{y_{010}|x\} + \mathbf{E}\{y_{010}|x\} - \mathbf{E}\{y_{000}|x\}$.

Substituting (65) into (64) gives a *difference-in-difference* (DID) formula for the conditional average effect of treatment on the treated,

(66) $\mathbf{E}(TT|x) = \mathbf{E}(y_{111}|x) - \mathbf{E}(y_{010}|x) - \mathbf{E}(y_{001}|x) + \mathbf{E}(y_{000}|x)$;

see Heckman and Robb (1986). If in particular, the determination of prices is described by the hedonic regression model (63), and the disturbance in this equation satisfies the critical assumption $E(\varepsilon_{wnt}|w=n\cdot t,n,t,x) = 0$, then one has

(67) $\delta_1 = \mathbf{E}(TT|x)$.

If the regression model (64) satisfies Gauss-Markov conditions, then the OLS estimator of δ_1 is an estimate of the average effect of treatment on the treated. The question then is what economic assumptions imply Gauss-Markov conditions, or other conditions under which alternative estimation methods yield consistent estimates of δ_1.

The additive linear specification of the regression (63) is apparently very restrictive, although by redefinition of x to include suitable transformations and interactions, and generalization of ε to allow conditional heteroskedasticity, it can be treated as a method of sieves approximation to a more general model. However, an approach introduced by Matzkin (2002) provides a more direct generalization of (66), and allows estimation of generalized moments of TT other than the conditional mean, an important consideration for policy applications. Describe the determination of y by a model

(68) $y = h(w,n,t,x,\varepsilon)$

in which no linearity or additivity conditions are imposed. Let $F_{nt}(y|x)$ denote the conditional distribution of y, given n, t, and x (with $w = n\cdot t$), $\omega_{nt}(x)$ denote the density of x given n,t, and $y = G_{nt}(q|x)$ denote the conditional quantile function; i.e.,

(69) $G_{nt}(q|x) = \sup\{y|F_{nt}(y|x) < q\}$.

The *effect of treatment on the treated* (TT) is defined, consistently with (64), is

(70) $TT(x,\epsilon) = h(1,1,1,x,\epsilon) - h(0,1,1,x,\epsilon)$,

the change in the outcome y at n = t = 1 when w changes from the counterfactual value of 0 to the factual value of 1, and x and ϵ are unchanged. The following result adapted from

Matzkin (2002) provides conditions under which the conditional expectation **E**{A(TT,x)|x} of any summable functional A(TT,x) can be identified and estimated.

Theorem 4. Suppose the model (68) is continuous in (x,ε), with x of dimension k, and satisfies

(i) h is strictly increasing in ε,

(ii) ε is statistically independent of (w,n,t,x),

(iii) h(0,1,1,x,ε) - h(0,1,0,x,ε) = h(0,0,1,x,ε) - h(0,0,0,x,ε); e.g., the effect of time on the treated in the counterfactual that they are untreated is the same as the actual effect of time on the untreated.

(iv) The densities ω_{nt} are positive at x.

Suppose TT satisfies (70), and A(c,x) is a measurable function of its arguments. Letting q denote a uniform random variable on (0,1), TT can be written

(71) $TT = DID(q,x) = G_{11}(q|x) - G_{10}(q|x) - G_{01}(q,x) + G_{00}(q,x),$

and if E{A(TT,x)|x} exists, then

(72) $E\{A(TT,x)|x\}= \int_0^1 A(DID(q,x),x)dq.$

Proof: Substituting (iii) into (70), one has

(73) $TT(x,\epsilon) = h(1,1,1,x,\epsilon) - h(0,0,1,x,\epsilon) - h(0,1,0,x,\epsilon) + h(0,0,0,x,\epsilon).$

In sub-population 11, ϵ has a CDF $q = \Gamma(\epsilon)$, and inverse function $\epsilon = \Gamma^{-1}(q)$. By (ii), it has the same distribution for the sub-populations 10, 01, 00. By (i), the conditional quantile functions $G_{nt}(q|x)$ in each of the sub-populations satisfy $G_{nt}(q|x) = h(n\cdot t,n,t,x,\Gamma^{-1}(q))$. Then, substituting $\epsilon = \Gamma^{-1}(q)$ in (73), one has the result that TT, written as a function of x and a uniform random variable q, has the form (71). Condition (iv) assures that TT is well-defined at x. Note that DID is not in general monotone in q. Finally, given A(c,x), one can substitute (71) to obtain (73). □

When A is linear, the expectation (72) will coincide with (66). More generally, (72) can be used as a basis for estimators of non-linear functions of TT, such as step functions that give the conditional CDF of TT. The next result gives a convenient estimator for (72) that is also adapted from Matzkin (2002).

Theorem 5. Suppose i.i.d. observations (y_i,n_i,t_i,x_i), for i = 1,...,N, and assume the observations are ordered so that $y_1 \le ... \le y_N$. Let $N_{nt} = \sum_{i=1}^{N} 1(n_i=n)\cdot 1(t_i=t)$ denote the number of observations in sub-population nt. Let ψ denote a continuous k-dimensional probability density that is positive in a neighborhood of zero, Λ denote a one-dimensional CDF with the property that Λ(z) - ½ has the same sign as z, and σ_N denote a window width that satisfies $\sigma_N \to 0$ and $N\sigma_N^{k+1} \to \infty$. Then the kernel estimator

$$(74) \qquad F_{nt}^{\#}(y|x) = (N_{nt}\,\sigma_N^{\,k+1})^{-1} \sum_{i=1}^{N} 1(n_i=n)\cdot 1(t_i=t)\psi((x-x_i)/\lambda_N)\Lambda((y-y_i)/\sigma_N)$$

satisfies $F_{nt}^{\#}(y|x) \to_p F_{nt}(y|x)$ at each x for which $\omega_{nt}(x) > 0$, and the conditional quantile estimator $G_{nt}^{\#}(q|x) = \max\{y|F_{nt}^{\#}(y|x)<q\}$ satisfies $G_{nt}^{\#}(q|x) \to_p G_{nt}(q|x)$.

Corollary 6. Suppose the hypotheses of Theorem 5, with $\Lambda(z) = 1(z\ge 0)$. Define

$$(75) \qquad Q_{nt}^{\#}(y|x) = (N_{nt}\lambda_N^{k+1})^{-1} \sum_{i=1}^{N} 1(n_i=n)\cdot 1(t_i=t)\psi((x-x_i)/\lambda_N)\cdot 1(y_i\le y).$$

Then,

$$(76) \qquad G_{nt}^{\#}(q|x) = \max\{y_i|Q_{nt}^{\#}(y_i|x) < q\}.$$

Proof: The estimator (74) from Theorem 5, with $\Lambda(z) = \mathbf{1}(z\geq 0)$, is

$$F_{nt}^{\#}(y|x) = (N_{nt}\lambda_N^{k+1})^{-1} \sum_{i=1}^{N} 1(n_i=n)\cdot 1(t_i=t)\psi((x-x_i)/\lambda_N)1(y_i\leq y) \equiv Q_{nt}^{\#}(y|x).$$

Then (76) follows from the inversion formula for the quantile function and the fact that $Q_{nt}^{\#}(y|x)$ is piecewise constant in y with knots at the observations y_i. □

The convenience of(76) is that it is unnecessary to invert $q = F_{nt}^{\#}(y|x)$ numerically. The estimator (76) is piecewise constant and right-continuous, with jumps at the knots $Q_{nt}^{\#}(y_i|x)$ for i = 1,...,N where $n_i = n$ and $t_i = t$. A continuous modification of (76) could be constructed by linear interpolation between the adjacent knots. Using the estimators $G_{nt}^{\#}(q|x)$ from Theorem 5 or Corollary 6 yields a computationally convenient consistent estimator of $\mathbf{E}\{A(TT,x)|x\}$.

Corollary 7. Suppose the hypotheses of Theorems 4 and 5. Define

(77) $DID^{\#}(q,x) = G_{11}^{\#}(q|x) - G_{10}^{\#}(q|x) - G_{01}^{\#}(q,x) + G_{00}^{\#}(q,x).$

Define an collection of points q_j for j = 0,...,4N+1, with $q_0 = 0$, $q_{4N+1} = 1$, and the remaining points given by $Q_{11}^{\#}(y_i|x)$, $Q_{10}^{\#}(y_i|x)$, $Q_{01}^{\#}(y_i|x)$, and $Q_{00}^{\#}(y_i|x)$ for i = 1,...,N, ordered so that $q_0 \leq q_1 \leq ... \leq q_{4N+1}$, . Then,

(78) $E^{\#}\{A(TT,x)|x\} = \sum_{j=1}^{4N} A(DID^{\#}(q_j,x),x)(q_{j+1} - q_{j-1})/2$

satisfies $E^{\#}\{A(TT,x)|x\} \to_p E\{A(TT,x)|x\}$.

With additional assumptions on the smoothness properties of the distributions $F_{nt}(y|x)$, densities ω_{nt}, the function A(z,x), and the kernels used in constructing the estimators in Theorem 5 and its corollaries, one can establish optimal rates for the window width λ_N, and establish asymptotic normality for the estimators (76)-(78).

The most serious questions about application of Theorems 4 and 5 to estimate pecuniary losses from an environmental injury are the validity of the assumption that the effect of time on the treated if left untreated is the same as the effect of time on the untreated, and the assumption that the distribution of the disturbance ϵ does not depend on n, t, and x. In the next section of this paper, we will examine the structure of hedonic equilibrium and its implications for the properties of the treated versus the untreated, and the properties of ϵ. At this point, it should be noted that ϵ summarizes the effects of unobservables that are primitive to the economy, including consumer preferences, unmeasured attributes of properties and neighborhoods, and imperfections in the functional specification of (68). Further, *selection* effects, in which consumer choice of neighborhood n induces a dependence of the distribution of ϵ on n, and f*eedback* effects, in which the causal link from n, t, and x is confounded by a reverse causal link from y to x, will make the DID estimator of TT inconsistent. Methods for dealing with selection and feedback effects can be found in Heckman and Robb (1986) and Altonji and Matzkin (2003).

We next explore in a very simple general equilibrium economy the validity of the fairly conventional "difference-in-difference" or "random effects hedonic regression" method for resource valuation. We show in two simple examples that this method does accurately measure the pecuniary loss from an environmental injury, but that it does not in general estimate WTP. In the first example, there is insufficient information from the market to identify WTP, or in other terms, to identify the distribution of preferences, because of bunching of consumers with heterogeneous infra-marginal preferences.. In the second example, where all consumers are at an active margin, it is possible at least in some cases to identify preferences and recover WTP from a modified measure of loss that does not in general coincide with the pecuniary loss. These examples do not construct a general equilibrium economy where the “difference-in-difference” fails to estimate the pecuniary loss accurately, but it is clear that one could do so, since this method removes confounding effects only under a special hypothesis, and it is simple to see that general equilibrium effects can easily violate this hypothesis.

Example 1

Consider an isolated city that contains two neighborhoods, n = 0 and n = 1. We will deal with two time periods, t = 0 and t = 1. The city contains N households, indexed i =

1,...,N, with N large. Every household is endowed with an amount y of a composite good, which we will call "income", received each time period. The price of the composite good is normalized to one. In addition, there is one consumer, household 0, who lives outside the city and is endowed with a stock of N_1 identical housing units located in neighborhood n = 1, fixed in t, and a reversible constant returns technology that can transform between housing units in n = 0 and income at a marginal cost r_t in period t. Let $\theta = N_1/N$ and assume $0 < \theta < 1$.

There is a possible environmental hazard, indicated by w = 1 if it is present, and w = 0 otherwise. At t = 0, there is no environmental hazard in either neighborhood, and at t = 1, the environmental hazard is present in neighborhood n = 1 but not in n = 0. Then $w = n \cdot t$ in the factual situation. We want to also consider the counterfactual situation with w = 0 at n = t = 1.

Summarizing, the economy has four commodities in each period: housing units in n = 0 and n = 1, an environmental hazard that may be present at t = 1 in n = 1, and a composite good. The environmental hazard is not directly priced or controlled by the market. Measured in units of the composite good, the rental price of a housing unit in state wnt is denoted r_{wnt}. In the factual situation, households face prices (r_{000},r_{010}) at t = 0 and prices (r_{001},r_{111}) at t = 1. In the counterfactual situation, households face prices (r_{001},r_{011}) at t = 1. There are no intertemporal markets, and no way to inventory the composite good.

Assume that the indirect utility of a household i = 1,...,N has the Gorman polar form

$$(79)\quad u = \mathrm{Max}\{y-r_{010}-\rho\lambda, y-r_{000}\} + \mathrm{Max}\{y-r_{111}-\rho(\lambda+1), y-r_{001}\}$$

in the factual situation, and the form

$$(80)\quad u = \mathrm{Max}\{y-r_{010}-\rho\lambda, I-r_{000}\} + \mathrm{Max}\{y-r_{011}-\rho\lambda, y-r_{001}\}$$

in the counterfactual situation. The term $\lambda > 0$ is the perceived net disamenity of n = 1 relative to n = 0 when there is no environmental hazard. The parameter ρ is random, with a cumulative distribution function $\Psi(\rho)$ that has a compact support contained in the non-negative real line. This reflects heterogeneity in the degree of aversion of households to the permanent features and the environmental hazard, when present, for n = 1. Despite the linear structure of indirect utility conditioned on neighborhood choice, the presence of this choice makes the random taste parameter ρ non-separable. Household 0, the capitalist, supplies M_t units in n = 0 at time t, and has an indirect utility function for state w

$$\text{(81)}\quad u = y_0 + N_1(r_{w11} + r_{010}) + \text{Min}(N\text{-}N_1,M_0)r_{000} - (M_0\text{-}N+N_1)r_0 + \text{Min}(N\text{-}N_1,M_1)r_{001} - (M_1\text{-} N+N_1)r_1$$

The capitalist will supply housing units in n = 0 only if $r_{00t} \geq r_0$, and will supply $N\text{-}N_1$ units at $p_{00t} = r_0$.

At t = 0, the consumers with ρ such that $\rho < (r_{010}\text{-}r_{000})/\lambda$ will choose a housing unit in n = 1. Then, $\theta = \Psi((r_{010}\text{-}r_{000})/\lambda)$. Then, the equilibrium is $r_{000} = r_0$ and $r_{010} = r_0 + \lambda\rho^*$, where $\rho^* \equiv \Psi^{-1}(\theta)$ is the threshold aversion level for residents of n = 1.

At t = 1, the consumers with $\rho(\lambda+1) < r_{111}\text{-}r_{001}$ will choose a housing unit in n = 1, implying $\theta = \Psi((r_{111}\text{-}r_{001})/(\lambda+1))$. Then,$r_{001} = r_1$ and $r_{111} = r_1 + (\lambda+1)\rho^*$.

There is also a counterfactual equilibrium with w = 0 for t = 1 and n = 1. In this case, $\theta = \Psi((r_{011}\text{-}r_{001})/\lambda)$. Then, $r_{001} = r_1$ and $r_{011} = r_1 + \lambda\,\rho^*$.

Comparing the equilibrium housing allocations, there is no switching between housing units across the various cases, as n = 1 will always contain the N_1 households that are least adverse to the disamenities of this neighborhood, whether or not the disamenity is amplified by the environmental hazard.

The sum of the utilities of the N consumers and the capitalist in state w is

$$\text{(82)}\quad U_w = 2Ny + y_0 - \mathbf{E}(\rho|\rho<\rho^*)(2\lambda+w)N_1.$$

Then, the compensating variation in income necessary to exactly offset the environmental hazard is

(83) $C^* = U_0 - U_1 = N_1\mathbf{E}(\rho|\rho<\rho^*) = \int_0^{\alpha*} N_1\rho\Psi'(\rho)d\rho/\Psi(\rho^*)$

$$= N_1\rho^* - \int_0^{\alpha*} N_1\,\Psi(\rho)d\rho/\Psi(\rho^*) \le N_1\,\rho^*.$$

The reason that this is not exactly $N_1\rho^*$ is that the infra-marginal households dwelling in n = 1 are less impacted by the additional disamenity of the hazard than the most adverse household at the extensive household.

The conventional difference-in-difference estimator of injury from the environmental hazard is

(84) $TT = (r_{111}-r_{010}-r_{001}+r_{000})N_1 = N_1\,\rho^*.$

Then, TT overstates C*. Obviously, this happens because TT attributes to every household in n = 1 the same loss as that of the marginal, most adverse household induced to stay in the neighborhood. The apparent reason that the usual consumer surplus calculation fails is precisely that the infra-marginal consumers have no dimension on which they can adjust to a margin, so their effect is like the effect of unsatisfied secondary market margins in general equilibrium which make first-order contributions to welfare calculations in second-best welfare analysis.

At least in the setup above, there is no information contained in the market prices on the distribution of the infra-marginal ρ's. Then, one cannot distinguish the extremes in which all the N_1 households in n = 1 have $\rho = \rho^*$, or N_1 - 1 households have ρ = 0 and one has $\rho = \rho^*$, so the true equivalent variation could be anything between ρ^* and $N\rho^*$. Because the market prices cannot distinguish these cases, they cannot identify the distribution of consumer preferences. This implies that to have any hope of identifying consumer preferences from hedonic market data, every consumer must be operating at active margins with dimensionality at least as great as the dimension of the preference space.

In this example, all housing units were identical, and the environmental hazard impacted utility in the same dimension as the "permanent" relative disamenity λ of n = 1. If households were also choosing among housing units on the basis of observable hedonic attributes x, some components of which might be continuous, could this eliminate bunching of infra-marginal consumers, restore a margin, and allow identification of preferences? If the permanent and hazard components of the neighborhood disamenities of n = 1 were assessed by households in separate dimensions, would this supply the needed margin? (Letting λ vary randomly across households would, mechanically, be equivalent to allowing preference variation in two dimensions.) The next example suggests that under the right conditions the answer is affirmative.

Example 2

A second example induces an active margin for every household affected by the environmental hazard. Assume now that within n = 1 at t = 1, housing units are differentiated by their distance m from the environmental hazard, and the disamenity declines with distance. Specifically, let G(m) be the CDF of distances from the hazard within n = 1, and assume that m is scaled so that the support is [0,1]. Assume that the permanent disamenity in n = 1 also declines with distance m. The price of housing in n = 1 then varies with m; write r_{w1tm} for the price function. Suppose that the utility of a housing unit at m in n = 1 for a household with preference parameter ρ is $u = I - r_{w1tm} - \rho(1-m)(\lambda + w)$. The utility of a housing unit in n = 0 is the same as before. Also, $\theta = N_1/N$ is as before the maximum share of households that can be located in n = 1.

Consider market equilibrium at t in state w. As before, $r_{00t} = r_t$. To avoid dealing with corner solutions where some dwellings nearest the hazard would be vacant at zero price, assume that the support of $\Psi(\rho)$ is an interval that includes zero. The price function will be increasing for m > 0. The price function will induce an allocation in which households with taste parameter ρ are located at $m = D_{wt}(\rho)$. Material balance requires $N\Psi(\rho) = N_1[G(D_{wt}(\rho))]$ for $\rho \leq \rho^*$, where $\Psi(\rho^*) = N_1/N$ is the proportion of the population occupying units in n = 1. Utility maximization requires that

$$(85)\qquad r_{w1tD(\rho)} + \rho(1-D(\rho))(\lambda+w) \leq r_{w1tm} + \rho(1-m)(\lambda+w).$$

The first-order condition for this optimization is

$$(86)\qquad 0 = \nabla_m r_{w1tm} - \rho(\lambda+w),$$

and the second-order condition is

(87) $0 < \nabla_{mm}r_{w1tm}$.

Then, the price function must be increasing and convex. Since at the boundary m = 1, the disamenities associated with n = 1 are zero, one has the boundary condition $r_{w1t1} = r_{00t} \equiv r_t$. Let $\Psi^{-1}(q)$ be the inverse function solving $q = \Psi(\rho)$. Then, the material balance condition implies $\rho = \Psi^{-1}(\theta G(m))$ in equilibrium, so that

(88) $\nabla_m r_{w1tm} = \Psi(\theta G(m))(\lambda+w)$.

Using the boundary conditions,

(89) $r_{w1tm} = \max\{0, r_t - (\lambda+w) \int_m^1 \Psi^{-1}(\theta G(s))ds\}$.

The consumers in n = 1 are the N_1 least adverse, and they do not switch units with w or t, remaining in the order of least adverse at m = 0 to most adverse at m = 1. One has $\rho^* = \Psi^{-1}(\theta)$ and $D_{wt}(\rho) = D(\rho)$, as the price change is just sufficient to induce consumers to stay in place. Computing the equivalent variation for the consumer with preference parameter ρ,

(90) $U_0(\rho) - U_1(\rho) = (I - r_{011D(\rho)} - \rho\lambda) - (I - r_{111D(\rho)} - \rho(\lambda+1)) = r_{111D(\rho)} - r_{011D(\rho)} + \rho$.

Integrated over the consumers in n = 1,

(91) $\Delta U = N_1 \int_0^{\alpha*} (r_{111D(\rho)} - r_{011D(\rho)} + \rho)\Psi(d\rho)$

$$= N_1 \int_0^1 (r_{111m} - r_{011Dm})G(dm) + N_1\mathbf{E}(\rho|\rho<\rho^*).$$

The capitalist incurs the utility loss $N_1\int_0^1(r_{111m} - r_{011Dm})G(dm)$, just offsetting the pecuniary gain to residents. Then, the net equivalent variation, giving the magnitude of the welfare loss, is

(92) $\text{WTP} = N_1\mathbf{E}(\rho|\rho<\rho^*) = N \int_0^{\alpha*} \rho\Psi'(\rho)d\rho = N \int_0^1 \Psi^{-1}(\theta G(s))\theta G'(s)ds.$

This is completely intuitive, since the transfers wash out and all that is left is the drop in utility ρ for each consumer in n = 1.

Now consider the conventional difference-in-difference estimator of injury,

(93) $TT = N_1\int_0^1(r_{111m}-r_{010m}-r_{001}+r_{000})G(dm)$

$$= -N_1 \int_0^1 \int_m^1 \Psi^{-1}(\theta G(s))dsG(dm) = -N_1 \int_0^1 G(s)\Psi^{-1}(\theta G(s))ds.$$

Compare this with $\text{WTP} = N \int_0^1 \Psi^{-1}(\theta G(s))\theta G'(s)ds = N_1 \int_0^1 \Psi^{-1}(\theta G(s))G'(s)ds$. This is not the same as -TT except in special cases.[26] If G′ is non-decreasing, then -TT < WTP.

In this example, the active margin allows the distribution of preferences to be identified for those in n = 1. Specifically,

---

[26] An example where WTP = -TT is $\Psi\rho = \rho$ and $G(m) = m^{1/2}$.

(94) $\nabla_m DID(m) = \nabla_m(r_{111m}-r_{010m}-r_{001}+r_{000}) = \nabla_m(r_{111m}-r_{010m}) = \Psi^{-1}(\theta G(m))$

allows recovery of $\Psi(\rho)$ for $\rho < \rho^*$. However, in n = 0 where there is no active margin, the distribution of preferences is not recoverable.

8. Conclusions

The purpose of this paper has been to illustrate the usefulness of Gorman's polar form as an ingredient in applied consumer welfare analysis, particularly in problems where consumer location in space is an endogenous consequence of consumers' reactions to changes in the economic environment. Just as a "parallel Engle curves" assumption across consumers in conventional consumer analysis allows demands to be represented as those of a single representative consumer, an analogous assumption on Engle curves across locations for consumers with heterogeneous tastes in space allows spatial choice probabilities to be represented as those of a "locationally representative" consumer. These Gorman preferences have sufficient latitude to accommodate quite general patterns of substitution among commodities, and at least locally reasonable Engle curves, and yet have sufficient structure to give tractable forms for compensating variations. These in turn permit one to study the impact on consumer welfare of changes associated with spatially distributed environmental effects, and to examine the question of whether there is sufficient information in market observations to identify consumers' WTP. With examples, we suggest that the key to identification is the presence of active extensive margins of sufficient dimensionality for each consumer.

References

Aasness, J.; Bjorn, E.; Skjerpen, T. (2003) "Distribution of Preferences and Measurement Errors in a Disaggregated Expenditure System," *Econometrics Journal*, 6, 373-399.

Altonji, J.; Matzkin, R. (2003) "Cross-section and Panel Data Estimators in Nonseparable Models with Endogenous Regressors," Northwestern University Working Paper.

Anderson, G.; Blundell, R. (1983) "Testing Restrictions in a Flexible Dynamic Demand System," *Review of Economic Studies*, 50, 397-410.

Anderson, S.; De Palma, A.; Thisse; J. (1988) "A Representative Consumer Theory for the Logit Model," *International Economic Review*, 29, 461-466.

Blackorby, C.; Boyce, R.; Russell, R. (1978) "Estimation of Demand Systems Generated by the Gorman Polar Form: A Generalization of the S-Branch Utility Tree," *Econometrica*, 46, 345-363.

Blackorby, C.; Primont, D.; Russell, R. (1978) *Duality, Separability, and Functional Structure*, North Holland.

Blackorby, C.; Shorrocks, A.; eds. (1995) *Separability and Aggregation: Collected Works of W.M. Gorman*, Vol. I," Oxford Clarendon Press.

Boiteaux, M. (1956) “Sur la gestion des monopoles publics astreints à l’équilibre budgétaire,” *Econometrica*, 24, 22-40.
Brown, D.; Matzkin, R. (1998) “Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand,” Cowles Foundation, Yale University, Discussion Paper: 1175.
Browning, M.; Meghir, C. (1991) “The Effects of Male and Female Labor Supply on Commodity Demands,” *Econometrica*, 59, 925-951.
Chipman, J. (1990); “Marshall's Consumer's Surplus in Modern Perspective,” in J. Whitaker, ed., *Centenary essays on Alfred Marshall*. Royal Economic Society, Cambridge University Press,78-292.
Chipman, J. (1994) ”Hicksian Welfare Economics,” in H. Hagemann; O. Hamouda, eds. *The legacy of Hicks: His contributions to economic analysis*, Routledge, 1994; 96-146.
Chipman, J. (2000) “The New Welfare Economics 1939-1974,” in T. Cowen, ed*., Economic Welfare*. Elgar Reference Collection. Critical Ideas in Economics, vol. 3. Cheltenham.
Chipman, J.; Moore, J. (1980) “Compensating Variation, Consumer’s Surplus, and Welfare,” *American Economic Review*, 70, 933-348.
Chipman, J.; Moore, J. (1990) “Acceptable Indicators of Welfare Change, Consumer's Surplus Analysis, and the Gorman Polar Form,” in D. McFadden and M. Richter, eds, *Preferences, uncertainty, and optimality: Essays in honor of Leonid Hurwicz*.. Westview Press; 68-120.
Dagsvik, J. (1994) “Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes,” Econometrica, 62, 1179-1205.
Dagsvik, J. (1995) “How Large is the Class of Extreme Value Models?” *Journal of Mathematical Psychology*, 39, 90-98.
Dagsvik, J.; Karlstrom, A. (2003) “Compensated Variation and Hicksian Choice Probabilities in Random Utility Models that are Non-linear in Income,” Statistics Norway working paper.
Dagsvik, J. (2002) “Discrete Choice in Continuous Time: Implications of an Intertemporal Version of the IIA Property,” Econometrica; 70, 817-31.
Daly, A.; Zachery, S. (1979) “Improved Multiple Choice Models,” in D. Hensher and Q. Dalvi, eds, Identifying and Measuring the Dimensions of Mode Choice, Teakfield.
Deaton, A.; Muellbauer, J. (1980) “An Almost Ideal Demand System,” *American Economic Review*, 70, 312-326.
Deaton, A.’ Muellbauer, J. (1980) *Economics and Consumer Behavior*, Cambridge University Press.
Diewert, W. (1974) “Applications of Duality Theory,” in M. Intriligator and D. Kendrick, eds, *Frontiers of Quantitative Economics*, II, North Holland.
Diewert, W. (1982) “Duality Approaches to Microeconomic Theory,” in K. Arrow; M. Intriligator, eds, *Handbook of Mathematical Economics*, 2, North Holland.
Dreze, J. (1964) “Some Postwar Contributions of French Economists to Theory and Public Policy: With Special Emphasis on Problems of Resource Allocation,” *The American Economic Review*, 54, Supplement, Surveys of Foreign Postwar Developments in Economic Thought, 1-64.
Dubin, J.; McFadden, D. (1984) “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345-362.
Fenchel, W. (1953) *Convex Cones, Sets, and Functions: Lecture Notes*, Princeton University mimeograph.
Gorman, W. (1953) “Community Preference Fields,” *Econometrica*, 21, 63-80.
Gorman, W. (1961) “On a Class of Preference Fields,” *Metroeconomica*, 13, 53-56.
Gorman, W. (1968a) “The Structure of Utility Functions,” *Review of Economic Studies*, 35, 367-390.
Gorman, W. (1968b) “Conditions for Additive Separability,” *Econometrica*, 36, 605-609.
Hausman, J. (1981) “Exact Consumer Surplus and Deadweight Loss,” *American Economic Review*, 71, 662-676.
Heckman, J; Robb, R. (1986) “Alternative Methods For Solving The Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes" in Howard Wainer, ed., *Drawing Inference From Self Selected Samples*, Springer-Verlag
Heckman, J.; Matzkin, R.; Neshelm, L. (2003a) “Simulation and Estimation of Hedonic Models,” in T. Kehoe, T. Srinivasan, and J. Whalley, eds., *Frontiers in Applied General Equilibrium*, Cambridge University Press.
Heckman, J.; Matzkin, R.; Neshelm, L. (2003b) “Estimation and Simulation of Nonadditive Hedonic Models,” NBER Working Paper 9895.
Hicks, J. (1939) *Value and Capital*, Oxford, Clarendon Press.
Honohan, P.; Neary, P. (2003) “W.M. Gorman,” CEPR working paper.
Hotelling, H. (1935) "Demand Functions with Limited Budgets", *Econometrica*, 3, 66-78.
Hotelling, H. (1938) "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", *Econometrica*., 6, 242-269.

Morey, E.; Sharma, V.; Karlstrom, A. (2003) “A Simple Method of Incorporating Income Effects into Logit and Nested-Logit Models: Theory and Application,” *American Journal-of Agricultural Economics*, 85, 248-53.
Jorgenson, D.; Lau, L.; Stoker, T. (1980) “Welfare Comparison under Exact Aggregation,” *American Economic Review*. 70, 268-72.
LaFrance, J.; T. Beatty; Pope, R.; Agnew, G. (2000) “U.S. Income Distribution and Gorman Engle Curves for Food,” *IIFET 2000 Proceedings*.
Lau, L. (1977) “Existence Conditions for Aggregate Demand Functions.” Technical report, Stanford University.
Manski, C. (2003) *Partial Identification of Probability Distributions*, Springer Series in Statistics.
Marshall, A. (1920) *Principles of Economics*, Macmillan.
Mas Colell, A.; Whinson, M.; Green, J. (1995) *Microeconomic Theory*, Oxford.
Matzkin, R. (2003) “Nonparametric Estimation of Nonadditive Random Functions,”*Econometrica*, 71,1339-75.
Matzkin, R. (2003) “Identification of Consumer Preferences when Individual’s Choices are Unobservable,” *Journal of Economic Theory*, forthcoming.
Matzkin, R. (2004) “Unobserved Instruments,” Northwestern University working paper.
McFadden, D. (1963) “Constant Elasticity of Substitution Production Functions," *Review of Economic Studies*, 30, 73-83,.
McFadden, D. (1966) “Cost, Revenue, and Profit Functions: A Cursory Review,” Lecture Notes, University of California, Berkeley.
McFadden, D. (1978a) "Cost, Revenue, and Profit Functions," in M. Fuss and D. McFadden (eds.), *Production Economics: a Dual Approach to Theory and Applications*, I, 2-109, North Holland.
McFadden, D. (1978b) “Convex Analysis,” in M. Fuss; D. McFadden, eds., *Production Economics*, Vol. 1, 383-408.
McFadden, D. (1978c) "Estimation Techniques for the Elasticity of Substitution and Other Production Parameters," in M. Fuss and D. McFadden (eds.), *Production Economics: a Dual Approach to Theory and Applications*, II, 73-124, North Holland.
McFadden, D. (1978d) “Modeling the Choice of Residential Location,” in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, 75-96, North Holland: Amsterdam, Reprinted in J. Quigley (ed.), *The Economics of Housing*, Vol. I, 531-552, Edward Elgar: London, 1997.
McFadden, D. (1981) "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, 198-272.
McFadden, D. (1999) “Computing Willingness to Pay in Random Utility Models,” in J. Moore, R. Riezman, and J. Melvin (eds.), *Trade, Theory and Econometrics: Essays in Honour of John S. Chipman*, Routledge.
McFadden, D.; Train, K. (2000) “Mixed MNL Models for Discrete Response,” *Journal of Applied Econometrics*, 15, 447-70.
McKenzie, L. (1957) “Demand Theory without a Utility Index,” *Review of Economic Studies*, 24, 185-189.
Muellbauer, J. (1976) “Community Preferences and the representative Consumer,” *Econometrica*, 44, 979-999.
Panzar, J.; Willig, R. (1976) “Vindication of a "Common Mistake" in Welfare Economics,” *Journal of Political Economy*, 84,1361-63.
Pollack, R. (1971) “Additive Utility Functions and Linear Engle Curves,” *Review of Economic Studies*, 38, 401-414.
Ramsey, F. (1927) “A Contribution to the Theory of Taxation,” *Economic Journal*, 37, 47-61.
Rockafellar, R. (1970) *Convex Analysis*, Princeton University Press.
Roy, R. (1942) *De l'utilite, contribution a la theorie des choix*, Paris, Hermann & cie.
Samuelson, P. (1947) *Foundations of economic analysis*, Harvard Univ. Press.
Shephard, R. (1953) *Theory of Cost and Production Functions*, Princeton Univ. Press.
Stone, J. (1954) “Linear Expenditure Systems and Demand Analysis,” *Economic Journal*, 64, 511-527.
Williams, H. (1977) “On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit,” *Environment Planning* A, 9, 285-344.
Willig, R. (1978) “Incremental Consumer's Surplus and Hedonic Price Adjustment,” *Journal of Economic Theory*, 17, 227-53.
Willig, R. (1978) “Consumer's Surplus without Apology: Reply,” *American Economic Review*; 69, 469-74.
Varian, H. (1992) *Microeconomic Analysis*, Norton.